# Multimodale datafusie voor tijdsruimtelijke analyse van het brandgedrag

# Multimodal Data Fusion for Spatio-Temporal Fire Behavior Analysis

Florian Vandecasteele

Promotoren: prof. dr. S. Verstockt, prof. dr. ir. B. Merci
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: computerwetenschappen

UNIVERSITEIT GENT

Vakgroep Elektronica en Informatiesystemen
Voorzitter: prof. dr. ir. K. De Bosschere
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2018 - 2019

ISBN 978-94-6355-193-9
NUR 984
Wettelijk depot: D/2019/10.500/1

# Examination board

**prof. dr. Steven Verstockt - promoter**
Faculty of Engineering and Architecture
Department of Electronics and Information Systems
Ghent University-imec

**prof. dr. ir. Bart Merci - promoter**
Faculty of Engineering and Architecture
Department of flow, heat and combustion mechanics
Ghent University

**prof. dr. ir. Filip De Turck - chair**
Faculty of Engineering and Architecture
Department of Information Technology
Ghent University

**prof. dr. ir. Sofie Van Hoecke - secretary**
Faculty of Engineering and Architecture
Department of Electronics and Information Systems
Ghent University-imec

**prof. dr. Nico Van De Weghe**
Faculty of Sciences
Department of Geography
Ghent University

**dr. Jaap De Vries**
VP Innovation, Science and Technology
FM Global

**dr. ir. Steven Puttemans**
Technologiecluster Elektrotechniek (ESAT)
KU Leuven

**ir. Xavier Deckers**
Senior Project Manager
Fire Engineered Solutions Ghent

# Acknowledgments

*When a man becomes a fireman his greatest act of bravery has been accomplished. What he does after that is all in the line of work.*

*Edward F. Crocker*

The choice to fulfill this thesis is closely linked to my personal interests and background. As a child I was already passionated about firefighting and this passion has only increased. When professor Verstockt and professor Merci asked me four years ago to join the former Multimedialab to work on the Prediction of turbulent Reactive Flows (PRETREF) project I never doubted. I am very thankful to have them as promoters for my research; their knowledge on computer vision and fire behavior is tremendous. Furthermore, they were well listening to all my needs, were very open to give feedback and finally they stayed always calm when difficult decisions needed to be made. Secondly, I am grateful to everyone that gave me valuable feedback and advice along the way, ranging from my direct colleagues up to all the people I met on conferences and seminars. Subsequently, I am very thankful to everyone that contributed to the real-fire experiments in the Craeybeck-tunnel, Campus POV and the VIPA/ WfrGhent facilities.

I am really admiring my current and former office colleagues in Ghent (i.e., Aza, Olivier, Baptist, Frederic, Jasper, Mi-Jung, Krishna, Tamir, Sander, Dieter, Kenzo, Samnang and Jelle), the colleagues in Kortrijk (i.e., Sofie, Johan, Benjamin and Jan), the people of the fire safety and engineering group, the people of the PRETREF project and all those that are not listed. Subsequently, I am also thankful to everyone responsible for the enjoyable lunch sessions (i.e., Glenn, Johan, Ruben, Vasileios, Niels, Tom, Kristof and Laura).

Furthermore, I would like to acknowledge my parents and parents in law that encouraged me to always look forward and that gave me the opportunity to study without any concerns. Finally, I would like to thank my tower of strength, Tine Dujardin. Without you I would not be the person that I am right now.

*Ghent, October 2018*
*Florian Vandecasteele*

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms

## A

| | |
|---|---|
| AEC | Architecture, Engineering and Construction |
| AI | Artificial Intelligence |

## B

| | |
|---|---|
| BCF | BIM Collaboration Format |
| BE-SAHF | Building Environment - Smoke Air Heat Flames |
| BIM | Building Information Modeling |

## C

| | |
|---|---|
| CAD | Computer Aided Design |
| CCTV | Closed Circuit Television |
| CFAST | Consolidated Model of Fire and Smoke Transport |
| CFD | Computational Fluid Dynamics |
| CNN | Convolutional Neural Network |
| CO | Carbon Oxide |
| COCO | Common Objects In Context |
| CSV | Comma-Separated Values |

## F

| | |
|---|---|
| FDS | Fire Dynamics Simulator |
| FFT | Fast Fourier Transform |
| fov | field of view |
| FPS | Frames Per Second |
| fpa | focal plane array |

## G

| | |
|---|---|
| GIS | Geographic Information System |
| GML | Geography Markup Language |
| GPU | Graphics Processing Unit |

## H

| | |
|---|---|
| HMD | Head Mounted Display |
| HOG | Histogram Of Gradients |
| HRR | Heat Release Rate |
| HSV | Hue, Saturation and Value |
| HVAC | Heat, Ventilation and Air Conditioning |

## I

| | |
|---|---|
| ICT | Information and Communication Technology |
| IDM | Information Delivery Manual |
| IFC | Industry Foundation Class |
| IFD | International Framework for Dictionaries |
| IoT | Internet of Things |

## J

| | |
|---|---|
| JSON | JavaScript Object Notation |

## L

| | |
|---|---|
| LOD | Level Of Detail |
| LOI | Level Of Information |

## M

| | |
|---|---|
| mAP | mean Average Precision |
| Moses | Mobile sensing for fire safety |
| MSE | Mean Squared Error |
| MVD | Model View Definition |

## N

| | |
|---|---|
| NLP | Natural Language Processing |
| NFPA | National Fire Protection Association |
| NIST | National Institute of Standards and Technology |

## O

| | |
|---|---|
| OGC | Open Geospatial Consortium |
| OSM | OpenStreetMap |
| OWL | Web Ontology Language |

## P

| | |
|---|---|
| PhD | Doctor of Philosophy |
| POV | Provinciaal Opleidingscentrum voor Veiligheidsdiensten |
| PRETREF | Prediction Of Turbulent Reactive Flows |
| PSAP | Product Specific Adoption Potential |
| PTZ | Pan, Tilt and Zoom |

## R

| | |
|---|---|
| RDF | Resource Description Framework |
| ReLU | Rectified Linear Unit |
| RGB | Red Green Blue |
| RGB-D | Red Green Blue - Depth |
| ROI | Regions Of Interest |
| RPN | Region Proposal Network |

## S

| | |
|---|---|
| SAD | Sum of Absolute Differences |
| SHC | Smoke and Heat Control |
| SLIC | Simple Linear Iterative Clustering |
| SSD | Sum of Squared Differences |
| SSD | Single Shot Detector |
| SUS | System Usability Scale |

| | |
|---|---|
| SVM | Support Vector Machine |
| SURF | Speeded Up Robust Features |

## T

| | |
|---|---|
| tf-idf | Term Frequency - Inverse Document Frequency |

## U

| | |
|---|---|
| URI | Uniform Resource Identifier |
| UV | Ultra-Violet |

## V

| | |
|---|---|
| VFD | Video Fire Detection |
| VGG | Visual Geometry Group |
| VIPA | Vlaams Infrastructuurfonds voor Persoonsgebonden Aangelegenheden |

## W

| | |
|---|---|
| WIFI | Wireless Fidelity |
| W3C | World Wide Web Consortium |

## X

| | |
|---|---|
| XML | Extensible Markup Language |

# Nederlandstalige samenvatting –Summary in Dutch–

Ondanks de toename aan preventiemaatregelen in alle landen van de wereld heeft een brand nog steeds een enorme impact op mensen, hun eigendom en het milieu. Volgens de 'Fire Safe Europe community'[1] worden er elke dag 5000 brandincidenten gemeld in Europa en in 90 procent van de gevallen gaat het om een gebouw of industriebrand. Geen enkele brand is echter dezelfde. Zowel subjectieve getuigenissen van de brandweer als objectieve resultaten van repetitieve testen onder dezelfde brandcondities staven deze hypothese. Voor standaard brandsituaties (zoals een brand in een kleine ruimte) is de variatie nog vrij beperkt en voldoen de bestaande procedures die uitgewerkt zijn op basis van het te verwachten brandverloop. Voor complexere branden (in onder andere ondergrondse parkings, industriegebouwen en atria) werken de standaardmethodes niet en kan het zelfs gevaarlijk zijn deze blindelings te volgen. Om toch een snelle en efficiënte interventie uit te voeren is extra informatie omtrent de omgeving, het gebouw en de brand noodzakelijk in het beslissingsproces.

De combinatie van geografische informatie, sensordata en gebouwinformatie wordt tot op heden weinig gebruikt bij interventies. Nochtans merken we een grote toename in video data (i.e, in 2006 werden er 10 miljoen camera's verkocht terwijl er voorspeld wordt door IHS Markit [2] dat er 130 miljoen toestellen zullen verkocht worden in 2018). Alsook is er een toename in GIS/map data en gebouwmodellen. Momenteel is er echter geen centraal noch decentraal beheersysteem waardoor deze data en modellen moeilijk toegankelijk zijn. Op dit moment bestaan er verder ook weinig tot geen systemen (op basis van video en/of andere sensordata) die een benadering kunnen geven van de huidige fase in het brandverloop of over de huidige structurele toestand van het brandende gebouw. Indien een brandweerploeg aankomt, moet de officier beslissingen nemen met beperkte informatie en onder sterke tijdsdruk (waardoor hij/zij niet over de mogelijkheid beschikt om verschillende sensor- en videostromen te analyseren). De locatie van de brand, het aantal slachtoffers en hun positie, alsook de structuur en de lay-out van het gebouw zijn echter onbekende factoren die wel de beslissingen omtrent de brandbestrijdingsstrategie kunnen beïnvloeden.

[1] https://firesafeeurope.eu/european-fire-safety-strategy-needed

[2] https://technology.ihs.com/598815

In deze thesis stellen we een methodologie voor die een antwoord biedt aan bovenstaand probleem. Het voorgestelde fireGIS raamwerk combineert geografische data en gebouwmodellen met sensordata zowel tijdens als na een incident. Met deze tool kan men een tijdruimtelijke analyse doen van het brandverloop om de ploegen tijdens de interventie bij te sturen. Tevens kan deze toepassing gebruikt worden om brandexperimenten objectief te evalueren. In het raamwerk worden de videobeelden (zowel thermische als visuele) geanalyseerd om de hoogte en dikte van de rooklaag en de locatie en dimensies van de vlammen te bepalen, alsook de zichtbaarheid in te schatten. Daarnaast zorgt een objectherkenningsmodule ervoor dat men een schatting kan maken van de totale brandlast en brandcurve in de ruimte. Door de object- en videodata te combineren en te visualiseren in de bestaande BIM gebouwmodellen heeft men een duidelijk overzicht van het brandverloop en wordt het situationeel bewustzijn verhoogd. Verder stelt deze thesis een methodologie voor om BIM en videobeelden te aligneren op elkaar met behulp van plaatsherkenning en semantische informatie. Op basis van deze mapping is het ook mogelijk om het BIM model te verifiëren aan de configuratie van de ruimte. Als voorbeeld: een ruimte kan in de loop van de tijd van bureaufunctie veranderen naar een opslagruimte, wat een andere brandlast symboliseert, maar zo kan ook het veranderen van een deuropening een groot effect hebben op het ventilatieprofiel en het finale brandverloop.

De bruikbaarheid van het voorgestelde raamwerk werd getest tijdens verschillende grootschalige brandexperimenten. In een eerste set van proeven werd het algoritme gebruikt om de doelmatigheid te evalueren van alternatieve brandveiligheidsmaatregelen in nieuwe zorgconcepten (grote gemeenschappelijke ruimtes). De software werd een tweede maal geëvalueerd tijdens reële brandproeven in de Craeybeckxtunnel. Deze proeven werden uitgevoerd in opdracht van het Agentschap Wegen en Verkeer van de Vlaamse overheid en hadden als doel het bestuderen van het effect van de ventilatie op de rookbeweging met behulp van de video datastromen. De grote meerwaarde van het voorgestelde raamwerk is de snelle en intuïtieve voorstelling van tijdruimtelijke sensor- en videodata. Hierdoor dient men niet meer manueel de grote set van videofragmenten te analyseren.

Om het voorgestelde raamwerk verder af te stemmen op de noden van de gebruikers werd een enquête gelanceerd binnen de Belgische brandweergemeenschap. Naast het gebruik van sensorwaarden was er een sterke vraag naar een tool die snel en eenvoudig de hoogtepunten uit een sequentie kan bepalen. Deze thesis stelt hiervoor een methodologie voor die keyframes selecteert, gelijkaardige beelden verwijdert en snelle semantisch zoekmechanismen combineert (bijvoorbeeld: toon alle beelden waarbij er nog mensen geëvacueerd worden).

Door de combinatie van data-science, video-en sensoranalyse, GIS/BIM, en brandonderzoek, heeft deze thesis een sterk multidisciplinair karakter en de resultaten ervan kunnen worden aangewend zowel op **academisch als professioneel vlak**. De volgende groepen kunnen gebruik maken van onze onderzoeksresultaten:

- **De brandweer** gebruikt de tijdruimtelijke sensordata, statische en dynamische videodata om interventies aan te sturen. De voorgestelde thermische objectdetectietool in combinatie met de BIM-data resulteert in een hulpmiddel om de effectieve brandlast in een gebouw in te schatten. Zo wordt het in de toekomst mogelijk om de slagkracht en de nodige hoeveelheid koelend vermogen objectief te berekenen.

- **De onderzoekers** gebruiken de sensor en BIM data als input voor de brandmodellen en CFD (computational fluid dynamics) modelleringen . Met behulp van onze analysetools wordt het mogelijk om accuratere, interactieve brandvoorspellingen te maken.

- **De architectuur- en studiebureaus** gebruiken de sensordata in combinatie met de BIM data voor preplanning en evacuatieanalyse. Door de visuele camerabeelden te gaan verifiëren aan het bestaande BIM model wordt het gebouwmodel aangepast aan de laatste versie.

- **De brandbeveiligingfirma's** zullen in de toekomst meer dan enkel branddetectie of rookdetectie vanuit een punt voorzien in een gebouw. De sensor- en videodata tools zullen samen gebruikt worden om meer inzicht te krijgen in de situatie en dit zowel in de periode *tijdens* als *na* het incident.

- **De overheidsinstanties** gebruiken het fireGIS raamwerk om een objectieve evaluatie te verkrijgen van reële brandexperimenten (zoals de studie uitgevoerd in de Craeybecktunnel en de VIPA studie naar brandbestrijdingmaatregelen voor ouderenvoorzieningen).

# English summary

Despite the increase in prevention measures in all countries around the world, fires still have a big impact on people, their property and the environment. According to the 'Fire Safe Europe community'[3] 5000 fire incidents are reported every day in Europe and in 90 percent of the cases these are building fires. The variety in fire development in these building fires is huge, i.e., each fire is different. Both subjective testimonies from the fire brigade and objective results from repetitive tests under the same fire conditions support this hypothesis. For standard fire situations (such as a fire in a small space), the variations are small and the existing procedures that have been worked out based on the expected fire development, work properly. For more complex fires (e.g., underground car parks, industrial buildings, atria) the standard methods do not work and it can even be dangerous to follow them blindly. Still, to ensure a fast and efficient intervention extra information about the environment, the building and its condition are necessary in the decision-making process.

Until now the combination of geographical information, sensor data and building information is only marginally used during a fire incident. However we notify a huge increase in video data (i.e, in 2006 10 million cameras were sold while it is predicted by IHS Markit [4] that 130 million devices will be sold in 2018). Furthermore, there is also an increase in the generation of GIS / map data and building models. At present, however, there is no centralized or decentralized management system, making the models difficult to access. Only a limited number of systems are able to give an indication of the current phase in the fire process or to give an indication of the current condition of the burning building. If a fire service crew arrives at the scene, the officer must make decisions under time pressure and with limited information. Due to the limited time, the officer does not have the ability to analyze all the different sensor and video streams. The location of the fire, the number of victims and their position as well as the structure and the layout of the building are unknown factors that, if known, can influence the firefighting strategy.

[3] https://firesafeeurope.eu/european-fire-safety-strategy-needed
[4] https://technology.ihs.com/598815

The proposed fireGIS framework combines geographic data and building models with sensor data. With this tool, a spatio-temporal analysis can be made of the fire process to adjust the teams during the intervention. This application can also be used to objectively evaluate fire experiments. In the framework the video images (both thermal and visual) are analyzed to determine the height and thickness of the smoke layer, to determine the location and the dimensions of the flames and to estimate the visibility. In addition, an object recognition module ensures that an estimate can be made of the total fire load in the room. By combining the object and video information and visualizing it in the existing BIM building models, decision makers have a clear overview of the fire progress and the situational awareness is increased. Furthermore, this thesis proposes a methodology to align BIM and video images using place recognition and semantic information. Based on this mapping, it is also possible to verify the BIM model with the configuration of the room. As an example: in the course of time, a room can change from desk function to a storage, which symbolizes a different fire load, but also changing a door opening can have a big effect on the ventilation profile and the final fire behavior.

The feasibility of the proposed framework was tested during several large-scale fire experiments. In the first set of tests, the algorithm was used to evaluate the effectiveness of alternative fire safety measures in new elderly care concepts (large common areas). The software was evaluated a second time during real fire tests in the Craeybeckx tunnel. These tests were carried out on behalf of the Agency for Roads and Traffic from the Flemish government and the goal was to study the effect of the ventilation system on the smoke movement with the help of video data streams. The added value of the proposed framework is its fast and intuitive representation of spatio-temporal sensor and video streams. Furthermore our platform avoids the manual analysis of large set of video clips.

A survey was launched within the Belgian fire service community to further align the proposed framework with the user-needs. In addition to the use of sensor values, there was a strong demand for a tool that would quickly and easily determine the highlights from a videosequence. In that context, this thesis proposes a methodology that selects keyframes, that removes similar frames and elaborates fast semantic search mechanisms (as an example, show all images where people are still evacuated).

Due to the combination of data-science, video and sensor analysis, GIS / BIM, and fire research, this thesis has a strong multidisciplinary character and its results can be used both in an academic and professional context. The following target groups can benefit from our research results:

- **The fire brigades** use the spatio-temporal sensor, static and dynamic video data to manage interventions. The proposed thermal object detection tool in combination with the BIM data provide a tool to assess the effective fire load in a building. This way, it will be possible in the future to calculate the necessary amount of cooling capacity.

- **The researchers** use the sensor and BIM data as input for the fire models and CFD (computational Fluid Dynamics) calculations. Using our analysis tools it is possible to make more accurate, interactive fire predictions.

- **The architecture and engineering study offices** combine the sensor data to use the BIM actively for pre-planning and evacuation analysis. Through the verification of the visual camera images, the building model is adapted to the latest version.

- **The fire protection firms** will, in the future, provide more than just point detection in a building. The sensor and video data tools will be used together to gain more insights into the situation and this both in the period *during* and *after* the incident.

- **The government agencies** use our tool to obtain an objective evaluation of real fire experiments (such as the study conducted in the Craeybeckx tunnel and the VIPA study).

# 1
# Introduction

*This chapter situates the importance of research on spatio-temporal fire characteristics estimation in enclosure fires. Secondly, this chapter provides an overview of the main research questions that are tackled in this PhD thesis. Thirdly, an overview of the national and international publications is given.*

## 1.1 Situating the importance of this research

Fires have a tremendous impact on people, their property and the environment in all the countries around the world. According to the Fire Safe Europe Community[1] 5000 fire incidents are reported each day in Europe. 90 Percent of these fires in the EU happen in buildings. Furthermore, the cost of fires is 126 billion euro each year and 4000 people are killed by a fire or its consequences every year. To tackle and reduce the impact and to increase the safety, fire and rescue services have created different firefighting strategies adapted to the stage of the fire development. In the design phase of the building, different fire development scenarios are taken into account and fire prevention services, architects, engineers and local authorities will work together to define evacuation routes, compartmentalization blocks, suppression and detection systems.

[1] `https://firesafeeurope.eu`

Since 2014 it is legally obliged for companies in Belgium to constitute an emergency plan. This should help the fire brigade and the workers to respond as quickly as possible during an incident. Different parts should be included in the file, such as a risk analysis, an evacuation procedure, the fire suppression and detection mechanisms and a clear plan of the building. Currently there is no standardization on the icons, colors or figures in an emergency file and each fire department can declare their own rules, the BIM standards could help to make a uniform system. Furthermore, when the building is finished, it is up to the users and the plant owner to update the initial information in case of changes. An automated system that takes videostreams into account could help detect construction modifications or room configuration changes (e.g., an office contained initially only chairs and tables and all of a sudden the room is a storage for boxes).

In recent buildings there are alarms panels indicating the activated detection and suppression regions. Still, this does not give information on the burning item or the detailed fire source location. The commanding officers will rely on their own experience, standard operation procedures and intuition to make decisions. Besides the smoke detection devices, in recent buildings there is an increased use of video footages, mainly for security and intrusion detection. The large number of videocameras could be exploited for fire incident management, for example, to give the status of the evacuation routes, to identify the amount of people inside the building, to estimate the smoke density or to define the fire fuel packages. Nevertheless, due to the extensive volume of data there is a need for an efficient mechanism to overview and process the large collections of sensor values and video footages.

Besides the visualization, the large set of sensor values can be exploited in a forecasting framework. Different initiatives are described in literature. Firegrid [1] proposed an integrated emergency response system where live sensor values (temperature or CO) are used to infer the incident conditions and to update the predictive models. Furthermore, a knowledge-based reasoning scheme is used on top of the predicted models to support the decision making process of an emergency responder. Beji et al. [2] created a two-zone fire forecasting system where the loss of physics was compensated by assimilating observed data (estimated HRR from detected flame dimensions) in the zone model. The inverse fire model problem is an optimization problem where you want to find the model invariants (the fire growth and the heat loss factor) that match the best with the observed values of the sensors. Jahn et al. [3] used the smoke detector and sprinkler activation time for the estimation of the fire characteristics (i.e., fire growth rate, fire origin location) by applying inverse Computational Fluid Dynamics (CFD) modeling.

Each of these methods are highly valuable for fire forecasting, but they still require manual detailed information of the potential fuel load, the building geometry and the ventilation openings. Although the forecasting models are potentially helpful to assist the decision process, there are four major reasons why the models are not yet incorporated in the standard fire fighting strategies:

1. Zone models, which are capable of estimating the fire properties faster than the advancement of the real-time incident, lack accuracy and reliability.

2. CFD models, which can include the detailed physics, are too slow (i.e., in the order of hours) to generate and they require manual validation and verification.

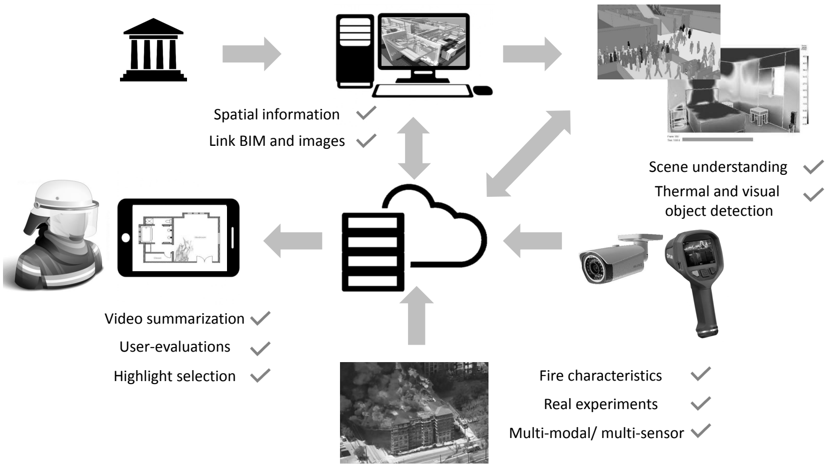
3. Accurate fire models require a detailed input of fuel packages, room configuration, fire progression, etc., and this data is currently not available.

4. Fire commanders are not always used to work with detailed simulation outputs. Due to the increased complexity and the possible overload of data this can even lead to wrong decisions.

In general there is a need for a system that clearly indicates the current phase in the fire process. A combination of geographic data and building models with sensor data both during and after an incident is considered by the fire fighting community to be the most suitable. Figure 1.1 shows the workflow of how we envision future fire fighting operations. The government facilitates the BIM data storage and from the BIM file, fire and evacuation models are calculated. The BIM data and the fire models are sent to a central system where there is additional input and feedback of the current smoke and fire behavior (from inside sensors). In addition, visual and thermal camera data is sent and processed by the central system. The camera and sensor input confirms or improves the latest BIM model in the system and the central system adapts the fire simulations. Finally, the commander in charge receives a structured overview of the building information, fire simulations and inside sensor values.

Different building blocks to realize a certain set-up were investigated in this PhD. Furthermore, the proposed model matches clearly with the Research road-map for smart fire fighting [4] proposed by NIST where they want to *"use the immense quantity of available data, the computational power to compute and communicate that data, the knowledge base and algorithms to most effectively process the data, convert it into significant knowledge/beneficial decision tools, and effectively communicate the information to those who need it - on the fire ground and elsewhere"*.

**Figure 1.1** The government facilitates the BIM data storage and from the BIM file, fire and evacuation models are calculated. The BIM data and the fire models are sent to a central system where there is additional input and feedback of the current smoke and fire behavior (from inside sensors). In addition, visual and thermal camera data is sent and processed by the central system. The camera and sensor input confirms or improves the latest BIM model in the system and the central system adapts the fire simulations. Finally, the commander in charge receives a structured overview of the building information, fire simulations and inside sensor values.




## 1.2 Spatial information: BIM and GIS

An important aspect in the proposed future fire incident management system is the use of GIS and BIM data. Geographic Information System (GIS) is a mechanism that captures, stores, analyzes and visualizes spatial, geographic information where the data is represented as real objects and geo-referenced to a global coordinate system. More specifically for fire fighters the GIS can, for example, display a toxic smoke plume or it can give valuable information about the water hydrants. Building Information Models (BIM) focus mainly on the indoor environment and configuration with high level of detail and with a rich set of spatial features and attributes referenced to a local coordinate system. GIS is commonly used for integrating, visualizing and analyzing information of a building while incorporating the context (i.e., environment, demography), while Building information models (BIM) and 3D/ CAD models are becoming more and more mainstream during the design and manufacturing phase of new buildings or during the renovation of ancient buildings. A 3D model with a connection to all the data and information about the complete building cycle allows the architecture, engineering and construction firms to handle the increased complexity of construction projects. Still there is a huge discrepancy between the original drawing, the latest design and the final 'as built' construction. Especially if we want to use the added value of BIM for fire forecasting or to assist firefighters during operation, verification is necessary between the BIM file and the real states. An open working space, for example, could be transformed into different separate rooms, or an originally empty room could be filled with boxes. To detect the actual room state we propose in Chapter 3 a deep learning based room configuration understanding mechanism.

## 1.3 Probabilistic room configuration understanding

With the recent advancements in deep learning object detection, image recognition and scene understanding are becoming mainstream. Over the past years there has been a significant gain in performance which allows real-time machine learning applications to be commercially used. (e.g., robotics, autonomous driving, surveillance and industrial monitoring). Despite the significant progress made in computer vision, machines are still not capable of achieving the same room understanding performance as a individual. Humans are trained to interpret the layout of a room (e.g., the position of the wall, ceiling and floor) and this even with a significant amount of clutter. People use the recognition and localization of objects in a particular scene to deduce the intention of a space (e.g., kitchen, living room) and a similar approach, incorporating contextual information, could help the existing deep learning and computer vision frameworks.

Previous computer vision research of IDLAB focused mainly on the flame and smoke detection [5, 6] with surveillance cameras and Passive Infrared based motion sensors. By analyzing the video and signals that the sensors generate it was shown that they could be used for the detection. However, with the increased use of CCTV (e.g., a report from the BSIA reported in 2014 that there are between 4 million and 5.9 million CCTV surveillance cameras in the UK) and handheld cameras, more detailed information about the scene and the specific smoke and fire properties is also available.

## 1.4 Spatio-temporal fire characteristics

An overview of the spatio-temporal changes of several fire indicators (smoke, visibility, fire dimensions) assists the commander to make appropriate tactical decisions. Currently, the firefighters are trained to manually identify certain fire indicators. This requires knowledge and understanding of fire dynamics, 'reading' and anticipating on rapid changing fire behavior indicators, skills in firefighting strategies and tactics. The **BE-SAHF** (Building, Environment, Smoke, Air Track, Heat, and Flame) organizing scheme, introduced by Grimwood et al. [7] is a well-known mechanism to assist in determining the stage of enclosure fire development[2], the location of the fire and suspected fire progress.

- The **building** and the construction in general will give an indication of the fire development. The building type will also indicate occupancy rate (e.g., an occupied school versus a domestic residence). Furthermore, the construction materials used in the building, the double-glazing and the insulation materials can hinder rapid fire growth due to the limited air supply or can, on the other hand lead to a rapid fire progress.

- The **environment** mainly focuses on the land topography, wind, humidity and extreme temperatures. A strong wind for example can accelerate the fire growth or can even induce a wind-driven-fire. Extremely low temperatures can decrease the buoyancy rate (i.e., upward force of the smoke) and the smoke can even form an inversion layer.

- The **smoke** features (i.e., the location, the volume, the thickness, the color, the buoyancy rate) will change the action plan as to how to fight the fire. Light, almost white, smoke can be due to the pyrolysis (highly combustible gases are formed of heated materials) or can be due to the presence of water vapor. In contrast, dark smoke can indicate a high amount of carbon released as soot in the smoke. Subsequently, the thickness and the spreading rate will affect the self-evacuation of victims.

- The **air** track can be visualized by the smoke pattern and it gives more information about the inlet and outlet openings. The air track indicators are velocity, turbulence, direction, and movement of the hot gas layer. If the air track is bi-directional (i.e., air in two opposite directions, fresh air going inside the building, hot smoke gases coming outside) at a certain opening, this may be the only adequate opening in the compartment.

[2]For more details and background of fire and enclosure dynamics we refer to Appendix A.3 at the end of this thesis.

- The **heat** will determine if it is still feasible to send firefighters inside. Subsequently, a sudden heat increase can be an indicator of an impending flashover or back draft.

- The final facet are the **flames** that are often the most visible indicator and described by a specific shape, movement, location and color. In a compartment fire for instance the flames will be yellow when the air supply is good, but when the air inlet and eventually the oxygen concentration is reduced, the flames will appear more red.

The interpretation of the fire characteristics and the usage of the BE-SAHF tool is currently depending on the knowledge and experience of the firefighters. A more objective, automated mechanism should be less error prone and could improve the spatio-temporal analysis. Finally, the automated calculation of fire characteristics could facilitate the updating of fire forecasting models as proposed in Section 1.1.

## 1.5 Fireground understanding and data visualization

In order to optimize the scene recognition and the fire characteristic analysis and to reduce the computational cost of subsequent video processing tasks, it is important to reduce the amount of video data by filtering out redundant and unnecessary frames, while preserving only those frames, distinctive and essential to capture the entire video content. Furthermore, providing the end-user, the fire commander, with a limited list of representative keyframes improves their exploration and search process.

Furthermore, visualizing too many sensor and data outputs to a first responder can cause overload difficulties. Due to an increased stress level it is possible to end up in a tunnel vision where the attention level is strongly narrowed. This can eventually lead to dangerous, deadly situations and a clear situational awareness strategy is necessary. Endsley et al. [8] defined situational awareness as the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future. There are several quantitative and qualitative selection criteria to select the correct and most valuable sensor output for fireground understanding. Within Chapter 5, questionnaires and automated visual outliers detection mechanisms are proposed to assist the decision process.

Finally, as stated by Hamins et al. [9] an improved fire ground understanding and in general a 'Smarter fire fighting' will possibly lead to following motivations:

- To save lives and minimize injuries to building occupants and community members due to fire.
- To improve fire fighter occupational health and safety.
- To enhance the operational efficiency of the fire service and the effectiveness of fire protection.
- To minimize property loss from fire.
- To minimize business interruption and loss of mission continuity due to fire.

## 1.6 Research questions

The previous sections described the lack of existing systems that combine spatial information from a BIM or GIS system, although it is already been illustrated that there is an added value for fire analysis and forecasting. Furthermore, there is a need for a mechanism that can give more insight information on the room configuration and the corresponding fire characteristics. Based on the problem statement we formulate the central question of this dissertation:

> **Can we develop a system to accurately detect, analyze and visualize spatio-temporal fire characteristics in enclosure fires and can we use the extracted information for fire behavior analysis and forecasting within a BIM framework?**

To facilitate answering this question, we break it down into smaller parts. Each chapter of this dissertation will focus on one of these parts. The first part of the research question is **the linking with a BIM framework**, described in Chapter 2. Research on building models for fire modeling are in their early phase. BIM for fire investigation can be decently used for fire safety design, or for post-fire analysis. Despite, the possibility of interactions, the building model is currently mostly used for visualization purposes making it a static source of information. Previous work has already investigated the possibilities to link BIM and low-cost point sensors for fireground understanding, but these approaches did not take into account the benefits of video analysis (large field of view compared to point sensors, robustness against dust and humidity, indoor scene analysis) and recent developments in semantics and feature learning research.

The second part of the central research question, i.e., **in enclosure fires** discussed in Chapter 3, requires a good knowledge of the room configuration, the scene type and subsequently an understanding of enclosure fire dynamics. Besides the object recognition, additional semantic information of the scene classification module can be used to further improve the localization accuracy. This enclosure fire task is different from previous research that combines satellite images for wildland fire forecasting.

The third aspect is the **accurate detection, analysis and visualization of spatio-temporal fire characteristics** which is considered in Chapter 4. A literature study revealed that the majority of sensor based fire analysis research is limited to triggering a fire alarm without further spatial or temporal analysis. Furthermore, for large scale fire experiments there is a need for automated processing of the large datasets. Several cameras are often monitoring the same scene (e.g., monitoring a tunnel) and the analysis is carried out manually without decent anomaly detection. In order to improve this, for instance, if one sensor is broken an alarm should be triggered and the inter correlated sensors (i.e., between the sensor groups, analysis of sensors in the neighborhood) and intra correlated sensors (i.e., inside the sensor group, analysis of the sensor over time) should be taken into account.

The fourth aspect of using **the extracted information for fire behavior analysis and forecasting** explained in Chapter 5 links the two facets. Firstly, the analysis links to the decision making and the situational awareness. Secondly, the forecasting assumes that the retrieved information can be incorporated into existing fire forecasting CFD or zone models.

## 1.7 Outline

The remainder of this thesis is organized as follows. Chapter 2 presents the global framework of building information models and the usability in a fire safety science context and fire incident management. Subsequently an overview is given on the combination of real-time sensing data and building information models. Finally we present some future research challenges for BIM enrichment.

Chapter 3 explores transfer learning techniques for visual and thermal object localization. Furthermore scene type classification (e.g., kitchen, bedroom) is used as input for probabilistic room configuration estimation. Finally, object tracking and 3D object understanding are discussed thoroughly.

Chapter 4 examines the spatio-temporal fire characteristics with the fireGIS framework and low-cost sensors. Firstly, an introduction is given into smoke reading and basic fireground understanding. Secondly, the real fire experiments are discussed more thoroughly. Subsequently, the smoke video analysis algorithm is discussed and the firemap generation algorithm is explained.

Chapter 5 proposes the fireground understanding and the data visualization. Firstly, the questionnaire that was launched in the Belgian firefighter community asking for the specific data needs and constraints, is explained. Secondly the video summarization framework for situational aware video retrieval is clarified. Thirdly, the fire characteristics visualization is discussed and finally, the visualization platforms and mechanisms are discussed more thoroughly.

Finally, chapter 6 lists the general conclusions of this thesis and points out future work for video fire analysis and BIM based forecasting.

Additional to the main chapters, two appendices are included in the thesis. Appendix A provides an overview of basic fire behavior and heat transfer concepts. Subsequently, appendix B provides an overview of basic machine learning and computer vision aspects. Both appendices can be consulted if the reader is not familiar with these techniques.

## 1.8 Publications

The research results obtained during this PhD research were published in scientific journals, presented at international conferences and shared in different articles in the fire safety science community. The following list provides an overview of the publications achieved during this PhD research.

### 1.8.1 International journal publications

**Vandecasteele Florian**, Bart Merci, and Steven Verstockt. ”Fireground location understanding by semantic linking of visual objects and building information models.” Fire Safety Journal, 91:10261034, 2017.

**Vandecasteele Florian**, Karel Vandenbroucke, Dimitri Schuurman, and Steven Verstockt. ”Spott : on-the-spot e-commerce for television using deep learning-based video analysis techniques”. ACM Transactions On Multimedia Computing Communications And Applications (TOMM), 13.3: 16-38, 2017.

**Vandecasteele Florian**, Bart Merci, and Steven Verstockt. ”Reasoning on multi-sensor geographic smoke spread data for fire development and risk analysis.” Fire Safety Journal, 86: 6574, 2016

### 1.8.2 National and international conference publications

**Vandecasteele Florian**, Bart Merci, Azarakhsh Jalalvand, and Steven Verstockt. ”Object localization in handheld thermal images for fireground understanding.” In Thermosense: Thermal Infrared Applications XXXIX vol. 10214. International Society for Optics and Photonics SPIE, 05, 2017.

Mario M Valero, Steven Verstockt, Oriol Rios, Elsa Pastor,**Vandecasteele Florian**, and Eulalia Planas. ”Flame filtering and perimeter localization of wildfires using aerial thermal imagery.” In Thermosense: Thermal Infrared Applications XXXIX vol. 10214. International Society for Optics and Photonics SPIE, 04, 2016.

**Vandecasteele Florian**, S. Verkaemer,Wouter Gevaert, and Steven Verstockt. ”Spatio-temporal data collection and visualization during firefighting incidents.” Research and advanced Technology in Fire Safety, 12, 2017.

Roeland TJampens, Francisco Hernandez, **Vandecasteele, Florian**, and Steven Verstockt. ”Automatic detection, tracking and counting of birds in marine video content.” Image Processing Theory Tools and Applications (IPTA), 1-6, 2016.

**Vandecasteele Florian**, Bart Merci, and Steven Verstockt. ”Smoke behaviour analysis with multi-view smoke spread data.” Interflam 2016. Interscience communications, 399408, 2016.

**Vandecasteele Florian**, Jeroen Vervaeke, Baptist Vandersmissen, Michel De Wachter, and Steven Verstockt. ”Spatio-temporal wardrobe generation of actors clothing in video content.” International Conference on Human-Computer Interaction. Springer, 448459, 2016.

**Vandecasteele Florian**, Bart Merci, and Steven Verstockt. ”Firegis : tijdruimtelijke videoanalyse van de brandverspreiding.” Fireforum Magazine, 5457, 2016.

**Vandecasteele Florian**, Esmee Vanbeselaere, Lore Vandemaele, Jelle Saldien and Steven Verstockt. ”Librarinth interactive game to explore the library of the future.” In 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, 9499, 2016.

**Vandecasteele Florian**, Tarek Beji, Bart Merci, and Steven Verstockt. ”Geographic reasoning on multi-modal fire spread data.” 2nd European Symposium on Fire Safety Science, 6368, 2015

# References


[1] Liangxiu Han, Stephen Potter, George Beckett, Gavin Pringle, Stephen Welch, Sung-Han Koo, Gerhard Wickler, Asif Usmani, José L Torero, and Austin Tate. *FireGrid: An e-infrastructure for next-generation emergency response support*. Journal of Parallel and Distributed Computing, 70(11):1128–1141, 2010.

[2] Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. *On the use of real-time video to forecast fire growth in enclosures*. Fire Technology, 50(4):1021–1040, 2014.

[3] Wolfram Jahn. *Using suppression and detection devices to steer CFD fire forecast simulations*. Fire Safety Journal, 91:284–290, 2017.

[4] Hernandez Paul. *Research Roadmap Traces the Path to'Smart'Fire Fighting*. 2015.

[5] Steven Verstockt, Sofie Van Hoecke, Tarek Beji, Bart Merci, Benedict Gouverneur, A Enis Cetin, Pieterjan De Potter, and Rik Van de Walle. *A multi-modal video analysis approach for car park fire detection*. Fire Safety Journal, 57:44–57, 2013.

[6] Steven Verstockt, Rik Van de Walle, Luis Alejandro Gonzalez Avila, Bart Merci, and John De Blonde. *Combining volume sensors with multi-modal video analysis for fire detection and forecasting*. In International Conference on Automatic Fire Detection, volume 2, pages 203–210, 2014.

[7] Paul Grimwood, Ed Hartin, and John McDonough. *3D Fire Fighting: Training, Techniques, and Tactics*. Fire Protection Publications, Oklahoma State University, 2005.

[8] Mica R Endsley. *Toward a theory of situation awareness in dynamic systems*. Human factors, 37(1):32–64, 1995.

[9] Anthony P Hamins, Nelson P Bryner, Albert W Jones, Galen H Koepke, et al. *Research Roadmap for Smart Fire Fighting*. Technical report, 2015.

# 2
# Spatial information

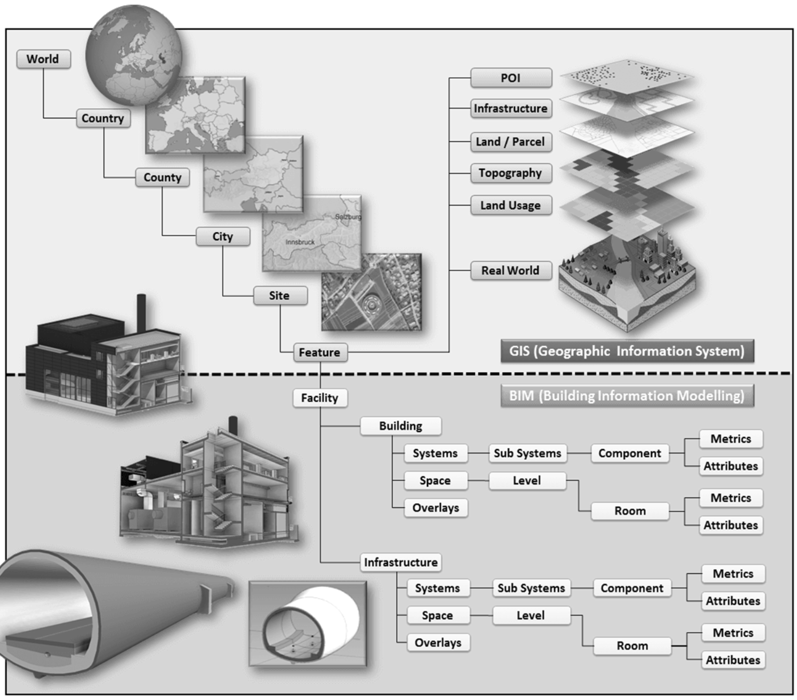
*This chapter focuses on the added value of Geographic Information Systems (GIS) and Building Information Models (BIM) in a fire safety science context. Initially, an introduction is given into the GIS and BIM basics and concepts, the Level Of Detail (LOD) and the Level Of Information (LOI). Furthermore, an overview is given on the current frameworks that combine geometric and topological information in a fire science context, as this is not available in literature. Additionally, modifications are proposed to the current architectures to increase the usability for regulators, designers and fire officers. Finally, more detailed information is given on how we can facilitate the image and BIM alignment from a conceptual computer vision perspective.*

## 2.1 Introduction

Undoubtedly smart cities are on the rise all over Europe and beyond [1]. A smart city is a region or city that uses different sensors in a collective network to connect citizens and the complete city infrastructure. The smart city projects being undertaken around the world all start from mapping existing building, utilities and transport infrastructure. Certain spatial information (included in BIM and GIS data) supports the decision making and facilitates the visualization and the interpretation of sensor data. CCTV camera positions and orientation, for example, can be registered in a BIM to easily find cameras with overlapping field of view or to find the nearby cameras. GIS mainly describes the current general environment state (i.e., pollution, traffic flow, energy consumption or gas and drinking water pipes) whereas BIM focuses on the detailed design and construction information of the building site. Figure 2.1 gives an overview of the data attributes specific for both models. GIS-attributes contain geographical and topographical information, ranging from the land, the region, up to the land-usage and the Points-Of-Interests. BIM is in the figure represented as the building with its components and their functionality. Furthermore detailed information is given about the metrics and the attributes.

Although the building models are a rich source of information it is important to remark that the majority of older buildings and sites lack a decent GIS or BIM model. Still early research results of Xue et al. [2] and Appolonia et al. [3] show that it is feasible to generate a simple BIM model from images and video footages. For more recent buildings, however, it becomes more and more standard to use BIM and GIS. An important side-note is that statistical results on the usage of BIM in the AEC community show that in 2011 only 13 percent of the population was aware of the term BIM whereas in 2018, 74 percent of the architects in the UK are currently using the framework (according to the NBS national BIM report [4]).

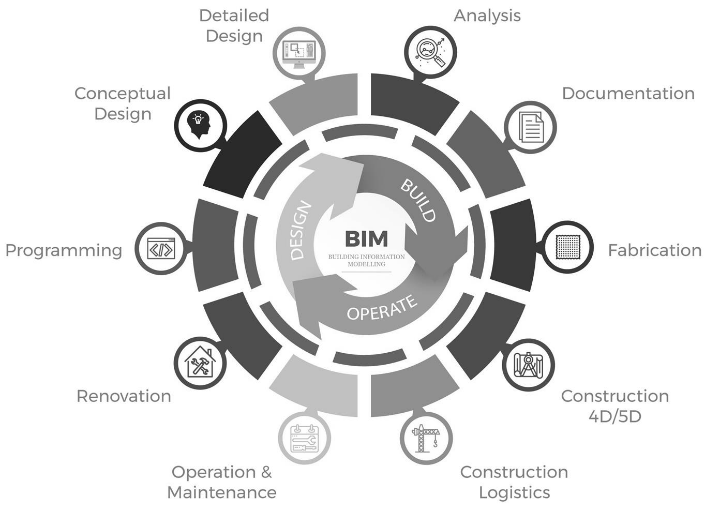
**Figure 2.1** Combining BIM and GIS information for building information management (source: https://www.ilf.com).




### 2.1.1 BIM

The definition of BIM has changed over the past decades. Initially, Charles Eastman defined it as a digital representation of the building process to facilitate the exchange and the interoperability of information. Lu et al. [5] added the need of a digital carrier. Eastman et al. [6] modified the definition of BIM, it should be more than the standard geometry. According to Eastman, the file should contain spatial relationships, geographic information, quantities and properties of the building components. Finally, Smith et al. [7] summarized BIM as a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle (defined as existing from earliest conception to demolition). Figure 2.2 gives an overview of the circular BIM process, starting in the earliest concept and design phase, up to the construction and the building phase and finally the operation and maintenance point.

**Figure 2.2** Generic circular BIM process (source: http://njhcadservices.co.uk)




BIM represents the building as models made of collections of building components such as walls, windows, doors and their relationship and properties. This is more than the standard geometry information that is stored in a classic 3D-CAD package (e.g., AutoCAD[1], SketchUp[2] or Blender[3]). Initially BIM was only intended to support the architectural, engineering and construction (AEC industries), but currently there are numerous applications in different fields (e.g. facility management, fire safety engineering, climate control). Furthermore, due to the international standards it is now possible to share and store information of other software vendors and to transform the BIM data into semantically[4] understandable data formats, such as EXPRESS, RDF, XML and OWL. Figure 2.3 gives an example of the different semantic outputs of a BIM snippet.

- EXPRESS [9] is an ISO defined data modeling language where the specifications are described in a strict scheme for a set of information requirements. Furthermore the format is plain text, which makes the language human readable.

[1] www.autodesk.be
[2] https://www.sketchup.com/
[3] https://www.blender.org/
[4] The semantic web is an extension of the current web in which information is given in well-defined meaning, better enabling computers and people to work in co-operation. [8]

- Resource Description Framework (RDF) [10] is a standard data interchange model in the world wide web consortium (w3c). The structured knowledge is often represented by triples, consisting of a subject, a predicate and an object. The relationships are indicated by an arrow and can eventually form a labeled graph. This can be compared with the structure of a sentence: the verb makes the connection between the subject and the object. For instance, the following statement: ”The building has a front door” describes the relationship between ”the building” (subject), ”has” (predicate) and ”front door” (object). Furthermore, in RDF, each triple element is uniquely identified by an Uniform Resource Identifier (URI). This allows unambiguous referencing and increases the interoperability.

- Extensible Markup Language (XML) is originally built to emphasize simplicity, generality, expandability and usability across different applications. This is done by storing the data in a plain text format without predefined tags. The main focus of the language is the presentation of the data, not the interpretation. A tag can contain detailed attributes and descriptions of the surrounded text, but compared to the RDF model it is not obligatory to connect a meaning to the tags.

```
<?xml version=1.0>
<house>
 <buildingelement> Column </buildingelement>
 <buildingelement> Beam </buildingelement>
</house>
```

- Web Ontology Language (OWL) [11] is built to develop ontologies (i.e., definitions and classification of concepts and entities and their relationship). The framework extends the capabilities of the RDF by adding more vocabulary for describing properties and classes (i.e., relationships between classes, equality, restrictions).

The main advantage of the OWL language is that topological relations between instances of these classes and some constraints on the usage of these classes and properties could be added to the code. Furthermore, intelligence can be added and new insights and correlations can be inferred from the data, without explitcit modelling.

**Figure 2.3** BIM snippet of the iGent tower in Ghent: EXPRESS code (top), RDF graph (middle) and OWL format (bottom).

```
ISO-10303-21;
HEADER;
FILE_DESCRIPTION(('ViewDefinition [CoordinationView]'),'2;1');
FILE_NAME('Project Number','2011-06-16T11:53:54',(''),(''),'Autodesk Revit Architecture 2011 - 1.0
FILE_SCHEMA(('IFC2X3'));
ENDSEC;
DATA;
#1=IFCORGANIZATION($,'Autodesk Revit Architecture 2011',$,$,$);
#2=IFCAPPLICATION(#1,'2011','Autodesk Revit Architecture 2011','Revit');
#4=IFCCARTESIANPOINT((0.,0.));
#5=IFCDIRECTION((1.,0.,0.));
#10=IFCDIRECTION((0.,0.,-1.));
#11=IFCDIRECTION((1.,0.));
#12=IFCDIRECTION((-1.,0.));
#13=IFCDIRECTION((0.,1.));
#14=IFCDIRECTION((0.,-1.));
#15=IFCSIUNIT(*,.LENGTHUNIT.,.MILLI.,.METRE.);
#16=IFCSIUNIT(*,.AREAUNIT.,.MILLI.,.SQUARE_METRE.);
#17=IFCSIUNIT(*,.VOLUMEUNIT.,.MILLI.,.CUBIC_METRE.);
#18=IFCSIUNIT(*,.PLANEANGLEUNIT.,$,.RADIAN.);
#19=IFCDIMENSIONALEXPONENTS(0,0,0,0,0,0,0);
#20=IFCMEASUREWITHUNIT(IFCRATIOMEASURE(0.01745329251994328),#18);
#21=IFCCONVERSIONBASEDUNIT(#19,.PLANEANGLEUNIT.,'DEGREE',#20);
#22=IFCSIUNIT(*,.TIMEUNIT.,$,.SECOND.);
#23=IFCUNITASSIGNMENT((#15,#16,#17,#21,#22));
#26=IFCAXIS2PLACEMENT3D(#3,$,$);
```

```
  <rdf:Description rdf:about="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcOrganization_1">
    <rdf:type rdf:resource="http://www.buildingsmart-tech.org/ifcOWL/IFC2X3_TC1#IfcOrganization"/>
    <ns0:name_IfcOrganization rdf:resource="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcLabel_417033"/>
  </rdf:Description>

  <ns0:IfcLabel rdf:about="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcLabel_417033">
    <ns1:hasString>Autodesk Revit Architecture 2011</ns1:hasString>
  </ns0:IfcLabel>

  <ns0:IfcApplication rdf:about="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcApplication_2">
    <ns0:applicationDeveloper_IfcApplication rdf:resource="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcOrganizat
ion_1"/>
    <ns0:version_IfcApplication>
      <ns0:IfcLabel rdf:about="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcLabel_417034">
        <ns1:hasString>2011</ns1:hasString>
      </ns0:IfcLabel>
    </ns0:version_IfcApplication>

    <ns0:applicationFullName_IfcApplication rdf:resource="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcLabel_4170
33"/>
    <ns0:applicationIdentifier_IfcApplication>
      <ns0:IfcIdentifier rdf:about="http://linkedbuildingdata.net/ifc/resources20170627_095352/IfcIdentifier_417035">
        <ns1:hasString>Revit</ns1:hasString>
      </ns0:IfcIdentifier>
    </ns0:applicationIdentifier_IfcApplication>
```

```
# baseURI: http://linkedbuildingdata.net/ifc/resources20170627_095352/
# imports: http://www.buildingsmart-tech.org/ifcOWL/IFC2X3_TC1

@prefix ifcowl:  <http://www.buildingsmart-tech.org/ifcOWL/IFC2X3_TC1#> .
@prefix inst:  <http://linkedbuildingdata.net/ifc/resources20170627_095352/> .
@prefix list:  <https://w3id.org/list#> .
@prefix express:  <https://w3id.org/express#> .
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .
@prefix owl:  <http://www.w3.org/2002/07/owl#> .

inst:   rdf:type     owl:Ontology ;
        owl:imports  ifcowl: .

inst:IfcOrganization_1
        rdf:type  ifcowl:IfcOrganization .

inst:IfcLabel_417033  rdf:type  ifcowl:IfcLabel ;
        express:hasString  "Autodesk Revit Architecture 2011" .

inst:IfcOrganization_1
        ifcowl:name_IfcOrganization  inst:IfcLabel_417033 .

inst:IfcApplication_2
        rdf:type  ifcowl:IfcApplication ;
        ifcowl:applicationDeveloper_IfcApplication  inst:IfcOrganization_1 .
```

In order to share and store the building information of other software vendors and to manage the BIM data, there is a clear need for a standard. The major standards in the open-source BIM community are defined by the BuildingSMART[5] community and are listed as follows:

- Information Delivery Manual (IDM): describes the building process and provides detailed specifications of the information that a user with a specific role would need to provide.
- Industry Foundation Class (IFC): sharing the information between software vendors and users with a large set of predefined classes. The underlying web technology allows building data to be easily linked to material, sensor and GIS data.
- BIM Collaboration Format (BCF): is an open standard XML schema to optimize the workflow between different software packages.
- International Framework for Dictionaries (IFD): ensures the mapping between different construction databases.
- Model View Definition (MVD): translates the process into technical requirements.

IFC ensures the fluent exchange of alphanumeric information attached to spaces, building elements and other components, between different platforms. The exchange of data, which is referred as the exchangeability, can be seen from two points, namely from the software side and from the building model. To ensure the exchange-ability of BIM elements in the IFC standard, the properties of each element are described in standardized Property sets (ifcPropertySet), similar to the URIs as described before.

[5] `https://www.buildingsmart.org/`

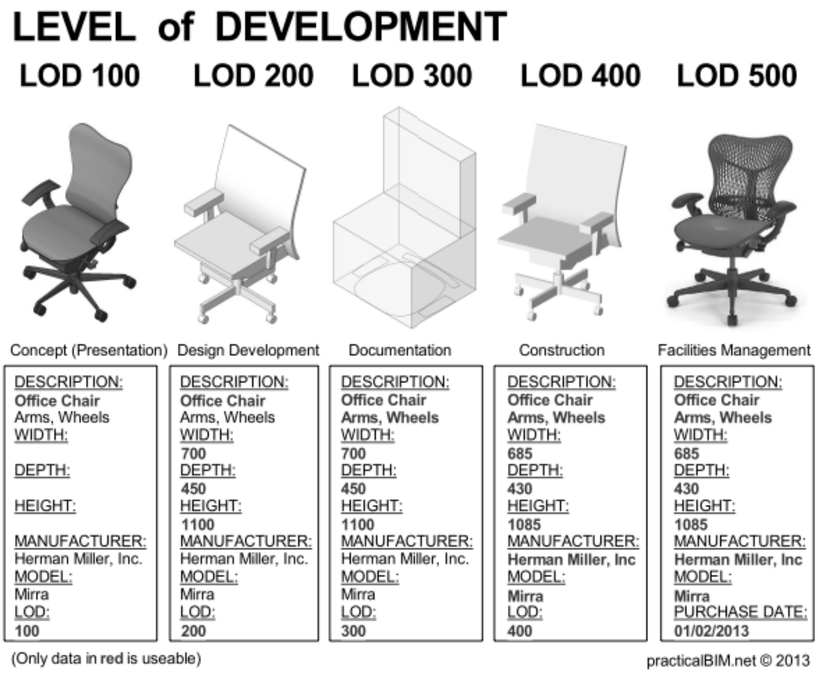
### 2.1.2 Level of development

Besides the standardization in property assets it is important to unambiguously describe the Level of Development (LOD) of the design. Currently, the definition of LOD is still discussed in the European Union. Mostly, the LOD comprises the graphic representation and the level of information. Within the BIM file a high level of information can refer to a file with specific factory sheets and maintenance instructions. In order to have a unified definition we currently refer to the LOD described by the Dutch national BIM platform [12]. Figure 2.4 gives an overview of the different detail levels:

- LOD 0: Simple spatial objects (i.e., spaces and volumes) with global dimensions and relations. Furthermore the function of the specific rooms can be described.
- LOD 100: Location and orientation, height, and volume of each level with its respective utility function.
- LOD 200: Material objects are modeled as generic building elements with global dimensions and properties.
- LOD 300: Spatial objects have exact dimensions and orientation; material objects are exact in quantity, dimension, shape, location and orientation.
- LOD 400: Material objects are highly accurate and precisely modeled. Furthermore details are given about the manufacturing and the construction of the objects on site.
- LOD 500: Objects are maximally detailed and all the information about shape, location, quantities and orientation is given.

In order to increase the usability of BIM data it is important to have a standardized information flow and a stimulated usage. The 'BIM loket' of the Netherlands [13] provides a central information point for BIM standards and it showcases BIM best practices. Such kind of initiatives are important to stimulate the effective and efficient use of the existing standards. A similar initiative should be launched in Belgium or Europe, as there is currently no data available. Furthermore, by increasing the usage of the BIM framework more and more construction and facility management applications will become available.

**Figure 2.4** Level of Development based on a chair, the red labels are the attributes present for that particular class. (source:www.practicalBIM.net)




### 2.1.3 GIS

Geographical Information Systems (GIS) are commonly used for the storage, extraction, modification, integration, visualization and analysis of spatial information. In literature following definitions are given:

- It serves for capturing, storing, analysis, and visualization of data that describe a part of the Earths surface, the technical and administrative entities, as well as findings of geoscience, economics, and ecological applications [14].

- It is an information system with a database of observables of spatially distributed objects, activities, or events, which can be described by points, lines, or surfaces [15].

Geopunt[6], for example, is a GIS platform with governmental Flemish web-services, which stores all the geographic government information (i.e., altitude, historical maps, living and society details). The web platform allows the stakeholders to achieve a very high level of integration due to the easy access to various data sources - it is even possible to upload your own datasets and analyze them on all kinds of map types. More specifically for fire fighters the GIS can, for example, display the calculated geographic spreading of a toxic smoke plume or it can give valuable information about position and condition of the water hydrants.

The combination of BIM and GIS has a high potential. GIS allows a decent visualization and modeling of the BIM project, whereas the BIM extends the GIS information with detailed building information. Furthermore, the end users have a better understanding of the impacts on the environment before, during and after the construction. GIS interacts with every stage of planning and development of a Smart City and facilitates an enormous spatially referenced database. For example, GIS can improve/ adapt the utilization of existing infrastructure capacity to improve the life quality or GIS can help to visualize spatial economic, environmental and social effects.

Currently there is no clear mapping between existing BIM and GIS frameworks. The datasets differ with respect to their semantics, geometry and level of detail. The indoor spatial information, for example, is standardized by the Open Geospatial Consortium (OGC) community in the indoorGML and cityGML application scheme whereas BIM is standardized on the IFC scheme. Due to the overlap in the features modeled in both domains as well as their differing strengths and weaknesses, it is widely acknowledged that the integration of data from both domains is beneficial and a crucial step forward for future 3D city modeling.

## 2.2 Spatial information in a fire science context

Geometric and topological information, such as the slope, the surface and the wind are commonly used to estimate wildfire growth. In enclosure fires, however, only limited spatial (meta) data of the building or room and its objects is used to prevent, forecast or tackle a fire. Still, in literature several initiatives explore the usability of building models and geographic data for analyzing fire science in enclosure areas and the following subsections a literature review will be given.

[6] https://www.geopunt.be/

### 2.2.1 Fire science context

Fire science is the study of all aspects related to fire, from fire behavior, fire safety design, fire forecasting, evacuation modeling, fire strategy and tactics up to fire investigation. Spatial information is a rich source of information to indicate the temporal state of the different factors that influence the development of the fire in a compartment (i.e., a confined space). A non-limiting list of these factors is given below:

- Properties of the fuel package (e.g., the type, amount, position, spacing and surface area),
- The geometry of the enclosure,
- The size, orientation and the location of the openings,
- The material properties of the enclosure boundaries,
- The room temperature and the temporal temperature change,
- The natural or mechanical wind speed, orientation and volume.

In Section 2.1.2 we explained the LOD. Starting from LOD300 there is detailed spatial information (BIM) of the objects (fuel packages), the room configuration and the generic building elements. This is perfectly in line with the list of factors that influence the fire development. In literature [16–18] different authors proposed ideas and solutions to combine BIM data for fire development analysis.

#### 2.2.1.1 Building information for fire safety design

Decent incident management starts with a good preparation of the fire safety design (and this already in the early design process). This is extensively discussed in literature. Zou et al. [16] gave an overview of the recent trends and possibilities for risk management during the building process (i.e., the safety management of the construction personnel). The main takeaway from this literature review is the use of the risk mitigation model. The model states that the risks should be identified and mitigated as early as possible in the design or planning phase of the building. However, this is not always possible due to time constraints, a lack of multi-disciplinary knowledge or due to ineffective communication. Furthermore, Zou et al. discussed the use of BIM for automated rule compliance checking. Fire safety requirements could be evaluated based on prescriptive rules (i.e., governmental rules and laws) or via a performance based design, where the BIM information is used as input for the simulation models. Finally, the conclusion of Zou et al. is that the BIM-based risk management is just emerging and there is currently no 'complete solution' available.

Jelenewicz et al. [18] discussed the use of BIM technology for general fire protection. The conflict checking of sprinkler system piping and equipments, the life safety drawings and information retrieval (e.g., egress paths, occupant load, wall rating) and the design of fire detection and suppression systems are discussed as possible use cases, but at the time of writing they are not integrated in the existing BIM packages. Finally, in their work some suggestions and features are proposed to integrate fire protection in BIM (i.e., automated sprinkler design, automated door configuration and selection, CFD and evacuation linking).

Wang et al. [19] present a solution to automate the evacuation route planning in two stages. Firstly, in the planning and design phase, the available safe egress time (CASET) and the required safe egress time (RSET) are calculated in a particular area. The algorithm will check if a particular route is acceptable according to the escape distance rules (specific for each region or country). Secondly, in the operating phase, a BIM 3D representation is made to help the users remember hazardous areas. Furthermore, safety equipment (e.g., fire extinguisher, fire hydrant) is marked in the web-based 3D model and for each equipment detailed information (i.e., brand, manufacturer, manuals, conditions) is stored and is easily accessible.

Zhang et al. [20] integrated semantic Natural Language Processing (NLP) for fully-automated building design checking. Evaluating the local or regional regulation codes is an important issue in fire safety design. In the work of Zhang, the BIM information is transformed into logic facts and the semantic reasoner is finally used to verify them against the textual codes. The main advantage of this approach is that the fire safety rules could be checked automatically without manual intervention. A major limitation, however, is that no specific validation is done for fire safety rules, which is our suggestion for their future work. Finally, it is important to remark that a decent fire safety design starts with a representative fire model and the building models could assist in this decision making process of the design fire (i.e., the fire load and the fire growth curve could be derived from the materials).

#### 2.2.1.2 Building models for fire investigation

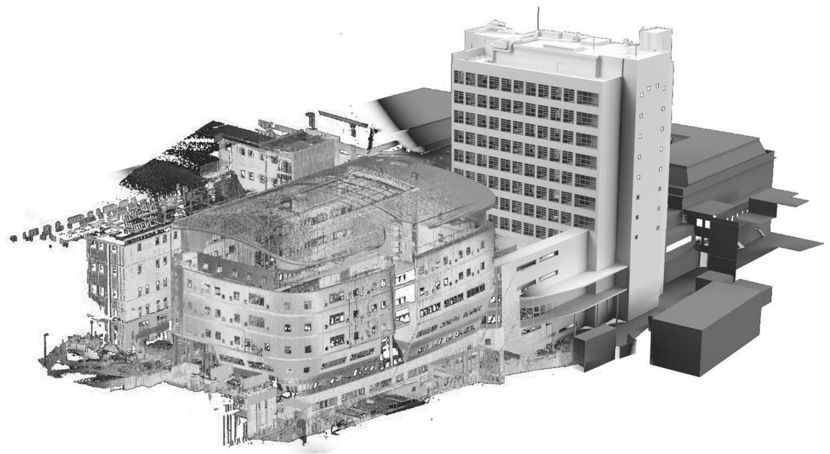
BIM can be used for performance based fire safety design (as mentioned in the previous section), or for post-fire analysis. Up till now, however, the building model is only used for visualization purposes. Wu et al. [21] proposed a system for data interchange between BIM and CFAST (Consolidated Model of Fire and Smoke Transport) [22]. After the simulation, simulation results are brought back into the BIM module for visualization. The most interesting block in their approach is the extraction of room configurations and compartment openings from building models. Furthermore, the forecasting with a fast zone-model and the visualization techniques are useful in our context. The disadvantage of their architecture, however, is that there is no verification of the current room configuration and they do not provide updates of the actual state of the building. The building model could be outdated and this could result in inaccurate simulation results. Furthermore, the fire source and location are not included in their work either. As such, their module is only useful for prevention purposes. By incorporating the actual, real-time sensing data, the system could be used for situational aware evacuation guiding and localization during a real fire. Lin et al. [23] on the other hand, used a 3D BIM-model for fire cause investigation and visualization. Their work only focused on presenting different kinds of data for fire disaster analysis and no combination/interaction is made between the sensor data and the building model. As such, the added value of this work is limited in our context.

Besides the data visualization it is necessary to have up-to-date building information for the fire modeling and the fire investigation. Franz et al. [24] used the point cloud information of a Google Tango mobile device to generate semi-automated geometric BIM models for post-fire investigation. Still point-cloud and complete 3D model generation are costly and an alternative could be that an existing building model in combination with real-time (visual or thermal) images are used to create the updated model (Figure 2.5 gives an artistic overview of the point-cloud scan to BIM process). The updating of an existing BIM model is one aspect we have studied in this thesis. For the thermal image analysis there is a link to the work of Lu et al. [25]. He showed that it is possible to get image-based localization with thermal images in reduced visibility settings. In Chapter 3 a profound explanation will be given on the object recognition and classification task for images based location understanding.

Wills et al. [26] map sensor data in a three-dimensional BIM for critical fire safety decision making. The most inspirational part of their methodology is the visualization of sensor data of real-fires in a 3D building model and the use of the sensing data for inverse fire model validation. The disadvantage of this approach, however, is that again there is no verification of the actual state of the building model and there is currently no possibility to integrate dynamic handheld sensors (e.g., sensor output of firefighters who are wearing temperature and pressure sensors).

Within the building model, different parameters are available to assist the fire investigation. Spearpoint et al. [27] examined the use and the interchangeability of IFC2x2 properties as a source of input data for zone modeling, ranging from building spaces and compartments, over fire suppression and detection systems to material properties (i.e., fire resistance, heat capacity or thermal conductivity). It is important to mention that these values could be used for fire investigation, but also

**Figure 2.5** Conversion from point-cloud data to BIM model West Sussex Hospital (source: scan2bim.info).



for incident management or evacuation planning. Currently, the IFC4 [28] is the new standard and, compared to the older standard,more consistency, connectivity between elements and incorporation of GIS elements are ensured.

#### 2.2.1.3 Building information for evacuation planning

Besides the use of BIM for fire analysis and forecasting there is also ongoing research on the use of BIM to support evacuation planning. Wang et al. [19], for example, used the 3D geometric BIM data and visualization results to support evacuation assessments and escape route planning. Currently, the objects in the building (e.g, furniture and fire safety equipment) are manually registered. With the computer vision based object recognition modules in our framework, an automated understanding and registration of the objects are made.

Chen et al. [29] used the BIM information to create an evacuation network graph. The indoor geometry is retrieved from the BIM model and a 'road' network graph is created with weights on its edges (for more detailed information and background about graph theory we refer to the work of West and Douglas et al. [30]). Secondly a routing algorithm is looking for the shortest path running the Dijkstra algorithm [31].

Zlatanova et al. [32] developed a similar system to Chen et al. for 3D model based emergency evacuation. The output of both systems could be used for evacuation guiding and routing decisions for rescue operations. The limitation of these systems is that there is no possibility to integrate sensing data during a fire. Congestion, hazardous regions due to smoke or fire could be highly valuable to adapt the evacuation or searching preferences.

Finally, Teo et al. [33] used the geospatial data from BIM and GIS for route planning in emergency evacuations, considering both the city and building scale. The automated generation of a semantic indoor network from the BIM/IFC files could be used for route planning (as proposed in the paper), but the semantic information can also be used for location estimation by applying a coarse-to-fine search approach (see Chapter 3 for more details).

### 2.2.2 Spatial information combined with real-time sensing

As indicated in the introduction it is important to have a decent visualization of the real-time sensor data. Wang et al. [34] proposed a conceptual web information service to combine the 3D model and the live sensing data. Their main idea is to use the interface for energy management, but the same mechanism can be used to visualize the sensor data of the temperature [35], the humidity [36] or the CO level during a fire incident. Furthermore, the increased use of Internet of Things (IoT) devices will automate the practical use of real aggregated data. IBM Watson IoT [37] and Honeywell[7], for example, show the combination of real sensors data to reduce energy costs. Subsequently, Kumer et al. [38] worked together with IBM on a proof-of-concept at Heatrow airport to test the feasibility of integrating data across the BIM asset to reduce the cost of unplanned maintenances.

Recently, more and more embedded or smart phone attachable sensors appear:

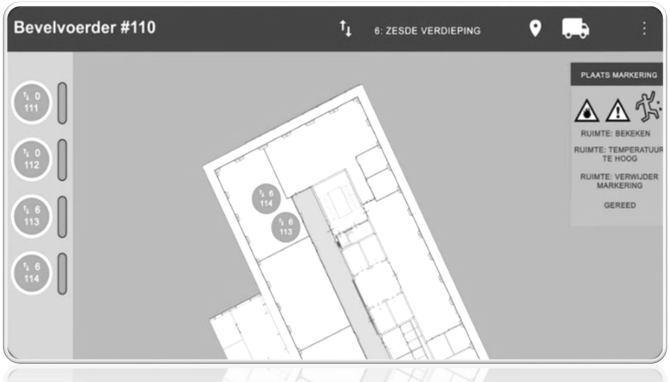
- Optical: single or dual spectral sensors (visual/ thermal), laser.
- Motion tracking: GPS, accelerometer, gyroscope, magnetometer.
- Environment analysis: humidity, UV, ambient light, temperature, pressure sensors.
- Sound and voice processing: multiple microphones (front-back).

The cost of these devices is decreasing and the amount of consumer applications is increasing. The majority of these systems, however, are still in an exploratory phase and are mostly not commercially available. Some research tracks explore the possibilities of these devices in case of a fire for real-time sensing, and the road map of NIST [39] gives some examples for further research on this topic. One example is the Moses (Mobile sensing for fire safety)[8] research project. This project

[7] https://www.honeywellprocess.com/en-US/pages/default.aspx
[8] http://techforfuture.nl/onderzoek/afgerond/mobile-sensing-safety-moses/

focuses on real-time situation awareness for fire fighters. The localization is estimated with foot mounted inertial sensors and a GPS sensor. This is further mapped on a 2D map of the building. The disadvantage of this approach once again is that the map needs to be up-to-date and that there is no feedback of the actual state of the building. Besides the indoor information it is valuable to have more detailed

**Figure 2.6** Moses project: tactical tablet visualization of the indoor position of the firefighters and their remaining oxygen level (source techforfuture.nl).




information of the condition of the firefighters. In that respect the Moses project uses probability calculations for the evaluation of personal and health sensing information, such as the heart rate and the breathing speed. Similarly, the heat stress risk could be calculated with wearable sensors and the core temperature [40]. A final interesting aspect of the Moses project is the tactical visualization of the data on a tablet (see Figure 2.6). The commander-in-chief gets an overview of the location of the firefighters on a 2D map. Future research could incorporate actual visibility estimations in the platform.

Another example of the use of real-time sensing is the work of Feese et al. [41]. They used smartphone data to monitor firefighter performance for post-fire analysis. The project focused on the monitoring of firefighters health during missions, monitoring of the environment of firefighters for toxic gases and high temperatures, and providing navigational support. Furthermore, the SmartRescue app from Radianti et al. [42] used smartphone sensors along with a Bayesian network classifier to assess the fire situation and to predict the further development, as well as to support an indoor positioning system. Within their application, the indoor localization is estimated by the strength of the WIFI signal. The limitation of this approach is that the system heavily relies on the WIFI signal which could be unavailable in case of a fire.

### 2.2.3 Commercial applications

Besides the conceptual and theoretical approaches different commercial packages are available that combine fire forecasting and simulation frameworks with BIM information. **Pyrosim** [43] is a graphical user interface for the FDS simulator [44] that helps to create and manage details of a complex fire. The building geometry and object assets are directly obtained from the BIM file. It is important that the BIM file is as detailed as possible. Small deviations in material types (i.e., heat insulation and condition properties) or object properties will directly affect the fire and smoke simulation results. The FDS fire simulator is then run based on the building input parameters, the selected fire source and the estimated fire growth. Finally, the visualization of the fire simulation results (i.e., the temperature, carbon monoxide concentration and the visibility loss) is done in the BIM model. **Cype**[9] is a commercial drawing package with several 'fire science' related packages. The 'CYPEFIRE design' part imports an existing BIM project and checks the accordance of the design with the prescriptive codes, such as the compartmentalization, evacuation distances and protection installations. 'CYPEFIRE sprinkler' is a tool to design the hydraulic fire protection network in an existing BIM drawing. Furthermore, the 'CYPECAD MEP' packages allows dynamic simulations of the evolution of fires in a designed building using the FDS and SmokeView tool. The fire load, the ventilation openings, the building elements, the sprinkler or detection mechanisms and the expected fire scenario are graphically selected and derived from the BIM file. Finally, the simulation results are brought back and visualized in the BIM package.

### 2.2.4 Future applications of building information models

Besides the usability for evacuation and fire analysis, the information retrieved by the BIM module could also give information to assess further development of the fire. For instance, the main material used for construction (wood, concrete, steel) will result in a different fire behavior and this could be deducted trough the BIM module. The dimensions of the different rooms could be used to calculate the necessary water capacity to extinguish the fire. The level of insulation could indicate if there is a higher risk of pyrolysis and flashover (see Appendix 1 for more details). The level of prevention services (compartmentalization (i.e., divide a structure in separated blocks to limit the fire and smoke spreading), sprinkler system (i.e., an active fire protection mechanism that automatically discharges water in case of a fire), mechanical smoke extraction systems (i.e., smoke outlets and vents to create a way to evacuate smoke and heat to open air)) can also be indicated in the module and facilitate the commander to make the appropriate decisions. Currently there

[9] `http://www.cype.com/en/`

is no module available that facilitates the communication/visualization of this information. Future research will perform user-based evaluations to see the best and easiest way to perform this task.

Another point to address in future research tracks is the spatial data reduction. Currently, the building models have a large amount of data and correlations, which complicates transfer and interpretation. It is not clear if the complete data interchange is always necessary. It could be possible to create a scalable BIM model, which can deliver the specific BIM data that is needed by the application. The ID-Lab group already has expertise in related challenges, such as hyperspectral data, genome and light-field compression [45]. Furthermore, the proposed techniques of IDLab allow data streaming and random access, and outperform the state of art compression results [46].

## 2.3 BIM enrichment and validation

As indicated earlier in this text, there are many existing buildings with incomplete or fragmented BIM data, which poses different challenges and problems for using the BIM in decision making processes. Especially in the firefighting case it is necessary to have up-to-date information of the building layout (see the BE-SAHF mnemonic used by the firefighters where the **B** stands for building (see Chapter 1)). More concrete, the inside structure, the indoor objects and the materials are influencing the strategic intervention decisions on the fire scene (e.g, a small change in the room configuration affects the fire progress[10] or a room with wooden walls will have another fire growth than a room with gypsum walls). In order to tackle these issues, we propose to enrich and validate the BIM data with information that we can extract from visual images.

To combine BIM data with image information, we first need to align both modalities. The second step is to generate additional semantic information, such as detecting the room type, object configuration and materials of which the objects and walls are made. To handle the alignment, a discussion is given in this section on the usage of point-cloud data for BIM enrichment. Secondly, in this section we explain the alignment problem and in Chapter 3, we give an overview of the mechanisms necessary to generate enriched semantic data.

[10] see Appendix 1.3

### 2.3.1 Point-cloud based BIM enrichment

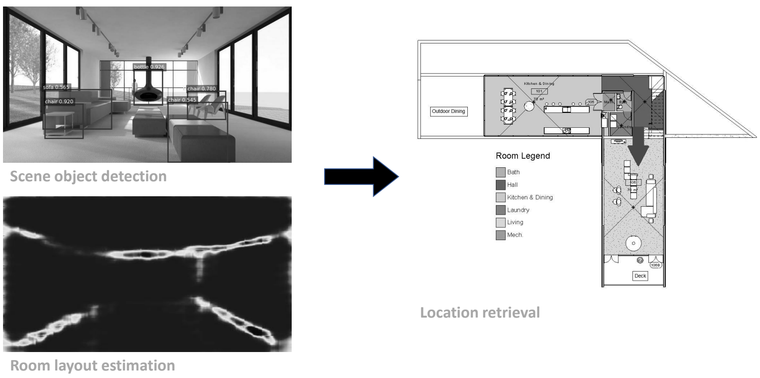
Patraucean et al. [47] provided a general overview of the as-built modeling process, mainly focusing on the generation of a detailed 3D model from a point-cloud file. Their goal is to create a semantically rich 3D model, composed of objects identified by geometry, relations and attributes. The first technique to create a detailed point-cloud model is to move (i.e., translate, rotate) the camera or laser scanner around in the building. In the literature this mechanism is referred to as Structure from Motion (SfM). A second technique is light detection and ranging (lidar) where a laser beam travels towards the area being scanned and back. By measuring the angles, the phase difference and the distances, accuracies from millimeters to centimeter could be achieved. Unfortunately, due to the complexity and the irregularity of shapes it is currently not possible to generate near-real time or real-time BIM models with this technique and the generation process still requires manual feedback, which makes the process unfeasible during an incident. The semi-automated BIM generation framework, however, is valuable to index 'older' buildings without a 3D plan or buildings with an outdated BIM file in advance. In the next subsection, we assume that a BIM file exist of a specific space or facility and we use the BIM file to estimate the position. Once the location in the building model is known, it is possible to update the model with fire-related sensor data (such as the visibility level and the smoke height) for visualization or for evacuation purposes. Furthermore, there could be an indication of the object that is burning and the fire properties of that object could be derived from the building model.

### 2.3.2 Image and BIM alignment

A first step in the BIM enrichment procedure is to align the current position (of the image-sensor) in the existing BIM model. Several commercial and academic frameworks are available to estimate the indoor location by means of a combination of a global positioning systems (GPS), an inertial measurement unit (IMU), a Doppler radar, a magnetometer, a compass, a pedometer or via wifi-fingerprinting. Kopsida et al. [48], for example, presented a markerless solution for the user-BIM alignment based on the Kinect Fusion for automated progress monitoring. This solution used the Kinect v2 sensor that acquires RGB-D data (i.e., the RGB color and depth image data) and the Kinect Fusion algorithm to reconstruct the existing scene and to estimate the pose. Li et al. [49] deployed an ad-hoc sensor network to visualize the indoor position in the BIM file, but no feedback was given on the exact orientation.

In this section two hypothetic mechanisms are proposed for image based location retrieval in the BIM file. The first mechanism is only suitable if the main indoor objects are described. In the second system we assume that the object location failed to estimate the correct location. In the current set-up, the existing sample project of the Revit package (i.e., A building information modeling software from Autodesk) is used as our information model and the data is transformed into a textual industry foundation class XML (ifcXML) file (see Figure 2.7 for the floorplan on the right side).

**Figure 2.7** Scene classification, object detection (upper left) and room layout estimation (down right) are used to estimate the position of the camera in the BIM model (right).




### Object based location retrieval

In order to extract the relevant Industry Foundation Class (IFC)-based spatial information, a first filtering is performed on the following elements in the ifc-file:

- **IfcBuilding** represents the building and is the main root of all the following and related elements.
- **IfcSpace** corresponds to a particular indoor partition (volume) constrained by walls. Furthermore, for each space the location, the direction and the polyline (a connected sequence of line segments created as a single object) coordinates are described. Additionally it is possible to define the purpose of the space. If the purpose is not described, but the indoor objects are tagged, it is still possible to derive the function of the room.
- **IfcDoor** defines inter-space relations and includes information of the location and the direction of the opening.

- **IfcWindow** is the description of a vertical and horizontal opening and includes information of the location, the material and the direction of the opening.

- **IfcFurnishingElement** is a generalization of all furniture related objects. The inherited values are the material type, the object class, the aggregation and the exact placement. The major known classes are semantically described by the buildingsmart alliance in the IFC4 standard [28] as follows:

  - CHAIR: Furniture for seating a single person.
  - TABLE: Furniture with a countertop for multiple people.
  - DESK: Furniture with a countertop and optional drawers for a single person.
  - BED: Furniture for sleeping.
  - FILECABINET: Furniture with sliding drawers for storing files.
  - SHELF: Furniture for storing books or other items.
  - SOFA: Furniture for seating multiple people.

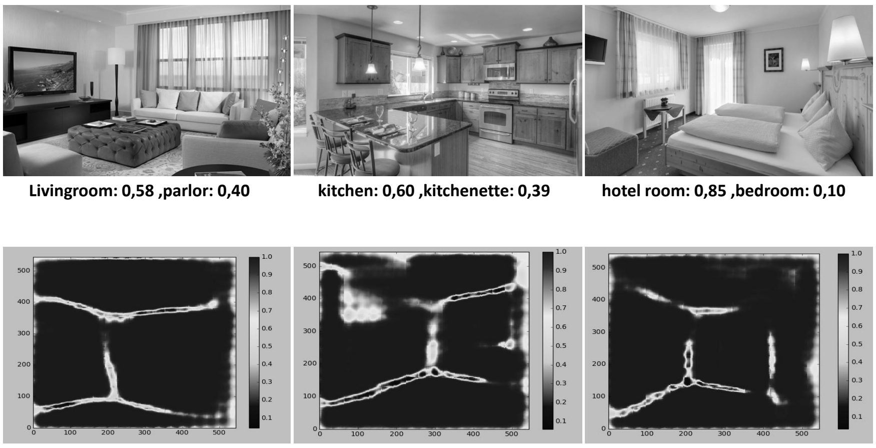
From the image (see Figure 2.7 left) the scene and object levels are extracted with computer vision techniques (more technical details and optimization techniques are described in Chapter 3). With the scene labels and their corresponding probability (parlor: 0.44, living: 0.17, lobby: 0.10, waiting room: 0.06) an initial search is performed on the BIM-based XML-file and, taking into account the spatio-temporal history of the localization, the room with the highest matching score is selected. Secondly, the objects (the classes that match with the IfcFurnishingElement description) are detected in the image (chair, sofa) and with these objects we perform a confirmation of the previous step and a rough estimate of the position of the imaging sensor in the room can be made.

#### Room layout location retrieval

In case the location is not sufficient a perspective alignment will facilitate the location retrieval in the BIM file. Firstly, the room layout detection mechanism will estimate the outline (i.e., the wall and floor corners) in the image. Secondly, the vanishing points are estimated in the image and in the particular BIM space. Finally, a RANSAC matching is performed on both vanishing points to find the best location based on the theory of Asadi et al. [50].

Room layout estimation focuses on estimating the 3D cuboid that defines the room, i.e., inferring the location and position of the walls and the ceiling. Existing solutions for layout estimation mostly rely on hand-crafted features and vanishing lines and they often fail in highly cluttered rooms. More interesting are the approaches of Ren et al. [51] and Delay et al. [52]. Ren et al., on the one hand, propose a coarse-to-fine strategy for indoor layout estimation. For a monocular input indoor image, a coarse layout estimation is generated with a multi-task fully convolutional neural network (MFCN). An example of this technique is shown in Figure 2.8, where the blue-green lines correspond to the most probable coarse layout estimation. In the second step, occluded lines and missing lines are filled and possible layout choices are ranked according to a predefined score function in the second stage. Delay et al., on the other hand, use a fully convolutional neural network (FCNN) in conjunction with a novel optimization framework for generating layout estimations.

**Figure 2.8** Room layout estimation, where blue and yellow colors correspond to the most probable outline region in the image.




Under perspective projection, parallel lines in 3D-space (e.g., wall lines) intersect in the image plane at vanishing points. From the layout estimation and the edges in the 2D image, the vanishing points are derived. Therefore, Rothers [53] algorithm is used to rank the edges and to determine the vanishing points. Furthermore, a similar procedure is performed in the BIM 3D image where for specific camera views a transformation is performed from a 3D image to a 2D view on it. As discussed in [50] it is important to transform the BIM coordinates to image coordinates taking the image camera parameters into account. Finally, the RANSAC iterating mechanism is used to find the best match in the BIM file using the vanishing points and the room layout.

## 2.4 Conclusions and future work

This chapter presented the basics of spatial information, more specifically the BIM and GIS framework. The first novelty of this research is the investigation and the exploration of spatial information specific for fire safety design, evacuation and fire investigation. The second novelty of this chapter is the semantic matching of computer vision techniques and BIM data to facilitate localization and situational awareness problems in fire emergency situations. It is important to remark that building models need to be up-to-date, complete and rich in detail. Currently, however, the building models are only used during construction and are mostly not updated afterwards, limiting their practical applicability. However, we expect that this will change in the upcoming years since more and more researchers start to explore the link with BIM in a wide range of application domains. The following chapter, focusing on probabilistic room understanding, can assist in the decision process and will help further automate the real-time configuration and layout detection in a building.

The results of the research presented in this chapter have led to new research initiatives and collaborations between IDLab, the geography department group of Ghent University and the department of flow, heat and combustion mechanics. Researchers of the geography department are developing an adaptive system for route instructions according to the perceived level of complexity at a decision point. The intricacy is derived from a space syntax mapped on the BIM visibility graph. Finally, the work presented in this chapter has inspired Chen et al. [54] to create a BIM-based real-time visualization and warning system for fire rescue services and Beata et al. [55] to develop a monitoring and visualization mechanism for post-ignition fire state analysis.

# References


[1] Andrea Caragliu, Chiara Del Bo, and Peter Nijkamp. *Smart cities in Europe*. Journal of urban technology, 18(2):65–82, 2011.

[2] Fan Xue, Weisheng Lu, and Ke Chen. *Automatic Generation of Semantically Rich As-Built Building Information Models Using 2D Images: A Derivative-Free Optimization Approach*. Computer-Aided Civil and Infrastructure Engineering, 2018.

[3] Fabrizio Ivan Apollonio, Marco Gaiani, and Zheng Sun. *A reality integrated BIM for architectural heritage conservation*. In Handbook of Research on Emerging Technologies for Architectural and Archaeological Heritage, pages 31–65. IGI Global, 2017.

[4] NBS (National Building Specification). *National BIM report 2015*. 2015.

[5] WW Lu and Heng Li. *Building information modeling and changing construction practices*. Automation in Construction, 2011.

[6] Chuck Eastman, Charles M Eastman, Paul Teicholz, Rafael Sacks, and Kathleen Liston. *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons, 2011.

[7] Peter Smith. *BIM & the 5D project cost manager*. Procedia-Social and Behavioral Sciences, 119:475–484, 2014.

[8] Tim Berners-Lee, James Hendler, and Ora Lassila. *The semantic web*. Scientific american, 284(5):34–43, 2001.

[9] Douglas A Schenck and Peter R Wilson. *Information modeling the EXPRESS way*. Oxford University Press, 1994.

[10] World Wide Web Consortium et al. *RDF 1.1 concepts and abstract syntax*. 2014.

[11] Deborah L McGuinness, Frank Van Harmelen, et al. *OWL web ontology language overview*. W3C recommendation, 10(10):2004, 2004.

[12] Aryani Ahmad Latiffi, Juliana Brahim, Suzila Mohd, and Mohamad Syazli Fathi. *Building information modeling (BIM): exploring level of development (LOD) in construction projects*. 2014.

[13] Ken Arroyo Ohori, Abdoulaye Diakité, Thomas Krijnen, Hugo Ledoux, and Jantien Stoter. *Processing BIM and GIS models in practice: experiences*

*and recommendations from a GeoBIM project in the Netherlands*. ISPRS International Journal of Geo-Information, 7(8):311, 2018.

[14] Norbert Bartelme. *Geoinformatik Modelle Strukturen Funktionen 4. Auflage*. Springer, 1995.

[15] Kenneth J Dueker. *Land resource information systems: a review of fifteen years experience*. Geo-Processing (Netherlands), 1979.

[16] Yang Zou, Arto Kiviniemi, and Stephen W Jones. *A review of risk management through BIM and BIM-related technologies*. Safety Science, 2016.

[17] Zhenyu Wang, Chunfeng Song, and Tao Chen. *Deep learning based monitoring of furnace combustion state and measurement of heat release rate*. Energy, 131:106 – 112, 2017.

[18] Chris Jelenewicz et al. *Building information modeling and fire protection*. In ASSE Professional Development Conference and Exposition. American Society of Safety Engineers, 2013.

[19] Shih-Hsu Wang, Wei-Chih Wang, Kun-Chi Wang, and Shih-Yu Shih. *Applying building information modeling to support fire safety management*. Automation in Construction, 59:158–167, 2015.

[20] Peng Zhou and Nora El-Gohary. *Ontology-based automated information extraction from building energy conservation codes*. Automation in Construction, 74:103–117, 2017.

[21] Chengde Wu, Saied Zarrinmehr, Mohammad Rahmani Asl, and Mark J Clayton. *Facilitating fire and smoke simulation using Building Information Modeling*. In International Conference on Computer-Aided Architectural Design Futures, pages 366–382. Springer, 2015.

[22] Richard D Peacock, W Jones, P Reneke, and G Forney. *CFAST–Consolidated Model of Fire Growth and Smoke Transport (Version 6) Users Guide*. NIST Special Publication, 1041, 2005.

[23] Lee-Kuo Lin, Yu-Huai Hsu, and Yu-Chen Hsiao. *A Study of Applying BIM Technique into Fire Disaster Investigation System*. In 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), pages 32–35. IEEE, 2016.

[24] Steffen Franz and Uwe Rüppel. *Supporting forensic fire cause analysis with next generation mobile devices by assembling a BIM based on multiple integrated sensor data*. In 22nd EG-ICE International Workshop, July 2015.

[25] Guoyu Lu, Yan Yan, Li Ren, Philip Saponaro, Nicu Sebe, and Chandra Kambhamettu. *Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging*. Neurocomputing, 173:83–92, 2016.

[26] Rosalie Faith Wills. *Development of a Cyber Physical System for Fire Safety*. PhD thesis, 2015.

[27] MJ Spearpoint. *Fire engineering properties in the IFC building product model and mapping to BRANZFIRE*. 2006.

[28] Thomas Liebich. *IFC4The new buildingSMART standard*. BuildingSMART Inter, 2013.

[29] Albert Y Chen and James C Chu. *TDVRP and BIM Integrated Approach for In-Building Emergency Rescue Routing*. Journal of Computing in Civil Engineering, page C4015003, 2015.

[30] Douglas Brent West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[31] F Benjamin Zhan. *Three fastest shortest path algorithms on real road networks: Data structures and procedures*. Journal of geographic information and decision analysis, 1(1):69–82, 1997.

[32] Sisi Zlatanova, George Vosselman, Kourosh Koshelham, Rob Peters, Bart Beers, Robert Voûte, Bart De Lathouwer, Matty Lakerveld, H Djurrema, and G Spaling. *Slimme 3D indoormodellen ter ondersteuning van crisismanagement in grote openbare gebouwen*. 2015.

[33] Tee-Ann Teo and Kuan-Hsun Cho. *BIM-oriented indoor network model for indoor and outdoor combined route planning*. Advanced Engineering Informatics, 30(3):268–282, 2016.

[34] Hongxia Wang, A Gluhak, and H Tafazolli. *Integration of BIM and live sensing information to monitor building energy performance*. In CIB W78 International Conference, Beijing, 2013.

[35] Jianli Chen, Tanyel Bulbul, John E Taylor, and Guney Olgun. *A case study of embedding real-time infrastructure sensor data to BIM*. In Construction Research Congress 2014: Construction in a Global Network, pages 269–278, 2014.

[36] Muhammad Arslan, Zainab Riaz, Adnan Khalid Kiani, and Salman Azhar. *Real-time environmental monitoring, visualization and notification system for construction H&S management*. Journal of Information Technology in Construction (ITcon), 19(4):72–91, 2014.

[37] Rob High. *The era of cognitive systems: An inside look at IBM Watson and how it works*. IBM Corporation, Redbooks, 2012.

[38] Bimal Kumar, Hubo Cai, and Makarand Hastak. *An Assessment of Benefits of Using BIM on an Infrastructure Project*. In International Conference on Sustainable Infrastructure 2017, pages 88–95, 2017.

[39] Hernandez Paul. *Research Roadmap Traces the Path to'Smart'Fire Fighting*. 2015.

[40] Mitra Baratchi, Lennart Teunissen, Peter Ebben, Wouter Teeuw, Jan Laarhuis, and Maarten van Steen. *Towards decisive garments for heat stress risk detection*. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pages 1095–1100. ACM, 2016.

[41] Sebastian Feese, Bert Arnrich, Gerhard Troster, Michael Burtscher, Bertolt Meyer, and Klaus Jonas. *CoenoFire: monitoring performance indicators of firefighters in real-world missions using smartphones*. In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pages 83–92. ACM, 2013.

[42] Jaziar Radianti, Mehdi Ben Lazreg, and Ole-Christoffer Granmo. *Fire simulation-based adaptation of SmartRescue App for serious game: Design, setup and user experience*. Engineering Applications of Artificial Intelligence, 46:312–325, 2015.

[43] ThunderHead Eng. *PyroSim User Manual*. The RJA Group Inc, Chicago, USA, 2011.

[44] Kevin B McGrattan, Randall J McDermott, Craig G Weinschenk, and Glenn P Forney. *Fire dynamics simulator, technical reference guide*. Technical report, 2013.

[45] A. J. Diaz-Honrubia, Johan De Praeter, J. L. Martinez, P. Cuenca, and Glenn Van Wallendael. *Reducing the complexity of a multiview H.264/AVC and HEVC hybrid architecture*. JOURNAL OF SIGNAL PROCESSING SYSTEMS FOR SIGNAL IMAGE AND VIDEO TECHNOLOGY, 90(2):249–258, 2018.

[46] Tom Paridaens, Glenn Van Wallendael, Wesley De Neve, and Peter Lambert. *AFRESh : an adaptive framework for compression of reads and assembled sequences with random access functionality*. BIOINFORMATICS, 33(10):1464–1472, 2017.

[47] Viorica Pătrăucean, Iro Armeni, Mohammad Nahangi, Jamie Yeung, Ioannis Brilakis, and Carl Haas. *State of research in automatic as-built modelling*. Advanced Engineering Informatics, 29(2):162–171, 2015.

[48] Marianna Kopsida and Ioannis Brilakis. *Markerless BIM registration for mobile augmented reality based inspection*. In Proceedings of the 16th International Conference on Computing in Civil and Building Engineering (ICCCBE2016), Osaka, Japan, pages 6–8, 2016.

[49] Nan Li, Burcin Becerik-Gerber, Bhaskar Krishnamachari, and Lucio Soibelman. *A BIM centered indoor localization algorithm to support building fire emergency response operations*. Automation in Construction, 42:78–89, 2014.

[50] Khashayar Asadi and Kevin Han. *Real-Time Image-to-BIM Registration Using Perspective Alignment for Automated Construction Monitoring*.

[51] Yuzhuo Ren, Chen Chen, Shangwen Li, and C-C Jay Kuo. *A Coarse-to-Fine Indoor Layout Estimation (CFILE) Method*. arXiv preprint arXiv:1607.00598, 2016.

[52] Saumitro Dasgupta, Kuan Fang, Kevin Chen, and Silvio Savarese. *DeLay: Robust Spatial Layout Estimation for Cluttered Indoor Scenes*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 616–624, 2016.

[53] Carsten Rother. *A new approach to vanishing point detection in architectural environments*. Image and Vision Computing, 20(9-10):647–655, 2002.

[54] Xiu-Shan Chen, Chi-Chang Liu, and I-Chen Wu. *A BIM-based visualization and warning system for fire rescue*. Advanced Engineering Informatics, 37:42–53, 2018.

[55] Paul A Beata, Ann E Jeffers, and Vineet R Kamat. *Real-Time Fire Monitoring and Visualization for the Post-Ignition Fire State in a Building*. Fire Technology, pages 1–33, 2018.

# 3

# A probabilistic method for room configuration understanding

*The previous chapter focused on the necessity of room and object information for fire behavior analysis. Within this chapter, our methodology for the indoor room configuration understanding process is explained. First, an introduction is given into computer vision principles, scene detection and visual object recognition. Subsequently, transfer learning is applied on a thermal image object dataset. Object detection in the thermal domain is a prerequisite to have a stable mechanism in case of reduced visibility in the visual domain due to smoke in the fireground. Next, because of the high computational cost of the object detection, methods for object tracking are proposed. Once an object is detected, we use the tracking to follow it in the subsequent frames. In order to further improve the accuracy of the room configuration understanding, this chapter also proposes a method to combine scene type classification and object detection. Finally some guidelines for detailed object understanding for fire forecasting are discussed.*

## 3.1 Introduction

The definition of computer vision has changed over the years from *"The construction of explicit, meaningful descriptions of physical objects from images", Ballard et al. [1]* up to *"Make useful decisions about real physical objects and scenes based on sensed images", Mahendran et al. [2].* The major change in definition is due to the inclined complexity in visual recognition tasks (e.g., face recognition, self-driving cars). Older studies heavily focused on feature engineering where features are generated based on statistical analysis, previous knowledge and feature performance evaluation. This requires expensive human labor and mostly relies on expert knowledge. Furthermore, the semantic gap between hand-crafted features and 'learned' features increases. A more recent trend is to use feature learning techniques where different positive and negative samples are shown to the system and, based on these examples, the parameters of the network are changed accordingly. The learned features are often more black box compared to the hand-crafted features (e.g., color or shape parameters). Still recent papers show that it is feasible to understand the deeper image representations [2].

*A feature is described in literature as a piece of information necessary to solve a computer vision task (e.g., an edge, a specific pattern or a descriptive color). To uniquely describe an image, features should be repeatable and accurate. More specifically the following conditions should be met:*

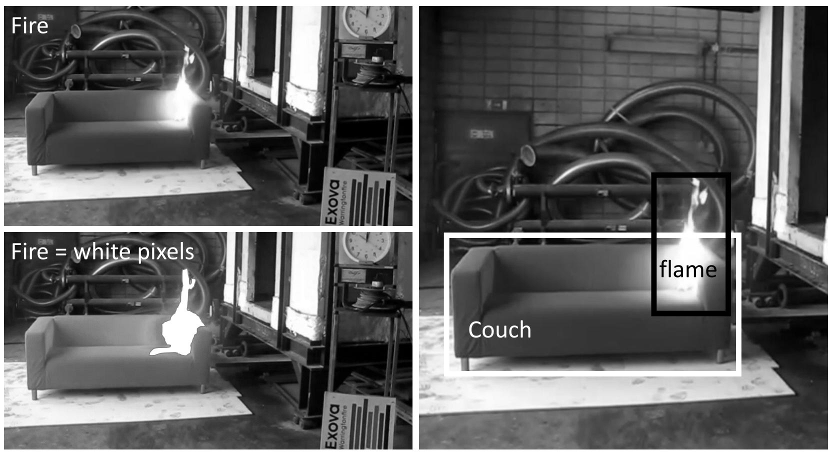
- *Invariant to translation, rotation and scale changes, i.e., the selected feature should be similar in case the camera is tilted, rotated or zoomed.*
- *Robust for covariant transformation, more specifically, the feature should be identical in case the viewpoint changes.*
- *Robust to lightning variations, noise, blur and quantization. The feature should, for example, remain stable all outdoor or indoor lightning conditions. The feature should also remain stable if the image is compressed (up to a certain level).*
- *Robust to occlusions, i.e., the feature should remain useful in case the object is only partly visible due to another object.*
- *Robust to inter-category variation. The feature should, for example, be stable to recognize a chair with different colors, different linings or patterns.*

Besides the definition, the core applications of computer vision have also changed over time. Initially, the main focus was on color and shape analysis, but currently, the key components are classification, localization and segmentation.

- Image classification: The general problem is to predict the category or label of an unseen image (e.g., image with fire or image without fire).
- Object localization: This task is bipartite and combines object detection (where is the object located in the image) and object recognition (which semantic class the object belongs to). The visualization can be, for example, a bounding box around the flames and a bounding box around the burning object.
- Image segmentation: This process indicates for each pixel in the image the semantic class. The result is a classification on fire pixel level instead of image level.

Figure 3.1 gives a conceptual overview of the components: classification, localization and segmentation in a fire context.

**Figure 3.1** Fire/non-fire classification (top left), object localization: couch and flame (right) and fire region segmentation (bottom left). (source: EXOVA-VIPA study)



## Global framework

In order to understand which object is burning, in which room the camera is located or to recognize the objects in the room, we propose the indoor-fire recognition framework shown in Figure 3.2. The input of the system is:

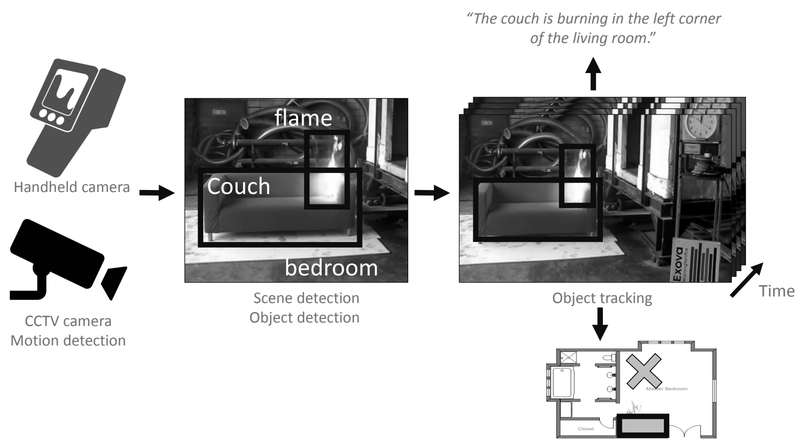
- A closed-circuit television (CCTV) camera, where motion detection notifies in which particular frame there are changes compared to the previous state.
- A monocular view of a hand-held thermal camera, where local motion detection is not useful due to the presence of global camera motion.

The scene detection and the object detection on the camera stream are combined to retrieve the burning object or the objects inside the particular room. Furthermore, surrounding object annotations will facilitate the interpretation of the fire behavior. By tracking the objects in time, we reduce the computational cost of the object detection block. However, it is important to note that new object detection mechanisms run at amazing speed on limited hardware. This could mean in the future that the tracking block could be neglected. Subsequently, if the burning object is known it can be linked to a building model of the specific location (see previous chapter). The smoke and fire sensor analysis is not taken into account here since it will be intensively discussed in the next chapter. Finally, the output of our system could be (for instance) the following text message: *The couch is burning in the left corner of the bedroom"* or *"The fire is located in the bottom right of the couch and the visibility is good*. The output of the systems can also be a visualization or update of the indoor BIM information. In order to achieve this final goal we focus on the following building blocks:

- A motion detection module,
- An object detection and recognition block,
- A scene classification block for location estimation verification,
- An object tracking framework,
- A smoke-fire location analysis block, which will be explained more thoroughly in next chapter.

The outline of this chapter is as follows: Section 3.2 will discuss the scene understanding and the object detection process in the visual domain. Subsequently, Section 3.3 will describe the object localization in the thermal domain. In the current set-up the motion estimation and object tracking are used as optimization tools to create a scalable system with multiple camera inputs. This will be further explored in Section 3.4 in combination with the network optimization. Furthermore, in Section 3.5 scene type classification and object detection are combined to

**Figure 3.2** Indoor-fire recognition framework: the input is a fixed or handheld camera, the processing consists of the scene and object detection on the camera stream, the object tracking and the generation of textual messages of the scene status or a visualization in the BIM file.




improve the probabilistic room configuration. Finally, 3D object recognition and material type detection are described to increase the object understanding process.

## 3.2 Visual scene understanding

Being able to classify the scene type (e.g., kitchen, living, salesroom) in the field of view of the camera will facilitate the location estimation and understanding. Due to differences in object types, shapes and texture features that construct the scene, this is still a challenging task. Recent deep learning based approaches [3], however, seem to be able to perform the classification with high accuracy. By embedding this semantic information of the scene in the room layout estimation, better estimations can be made in the localization process. As such, when the scene type can be detected, it is included in this process.

### 3.2.1 CNN models

Before going into detail in the classification optimization task a short overview is given of the state-of-the-art convolutional models. Some of the most popular architectures, that are used in many application domains, are VGG/ InceptionNet and Resnet. For more clarification of the terms and mathematical explanations, we refer to Appendix 1.

1. **VGG** - The VGG16 and the VGG19 network architecture were introduced by Simonyan et al. [3]. These network are characterized by their simplicity, using only 3x3 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling (i.e., a sample-based discretization process. It filters the less important features, passing the most important to the following convolutional layer.). Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier. The 16 and 19 stands for the number of weight layers in the network. There are two major drawbacks of the VGGNet, that makes deploying VGG a heavy task:

   - It is extremely slow to train compared to smaller networks. *The VGG-16 network takes 128.62ms for one step on a GTX 1080 GPUU device with 8GB of memory compared to the Alexnet that only takes 14.56ms on the same set-up [4].*
   - The network architecture weights themselves are quite large (in terms of disk/bandwidth). Due to its depth and number of fully-connected nodes, VGG is over 533MB for VGG16 and 574MB for VGG19. However, recent microSD cards of a size of 16GB are not uncommon anymore.

   VGG is used in many deep learning image classification problems. However, smaller network architectures are often more desirable, but require more optimization.

2. **ResNet** - Unlike traditional sequential network architectures such as AlexNet [5], OverFeat [6], and VGG [3], residual networks, such as ResNet [7], rely on micro-architecture modules to construct the network. The core idea of the micro residual module is the identity shortcut connection that skips one or more layers.

   First introduced by He et al. in 2015 [7], the ResNet architecture has become a seminal work, demonstrating that extremely deep networks can be trained using standard stochastic gradient descent (and a reasonable initialization function) through the use of residual modules. In 2016 He et al. updated their implementation by using identify mappings and this increased the general accuracy. Even though ResNet is much deeper than VGG16 and VGG19, the model size is actually substantially smaller (e.g., 102MB for ResNet50).

3. **Inception** - The Inception micro-architecture was first introduced by Szegedy et al. [8] in 2014. The goal of the inception module is to act as a multi-level feature extractor by computing 1x1, 3x3, and 5x5 convolutions within the same module of the network. The output of these filters are then stacked along the channel dimension before they are fed into the next layer in the network. Originally, this architecture was called GoogLeNet, but subsequent versions have simply been called Inception vN where N refers to the version number put out by Google. The weights for Inception V3 are smaller compared to the VGG and ResNet, summarized in 96MB.

In summary, each of the three architectures are suitable to be easily fine-tuned on new tasks and a trade-off should be made between accuracy and performance. Still, fine tuning for classification, localization or segmentation requires real-world annotated data (e.g., indoor fire scene datasets), which are currently not common available. Future work should in that context create a system where all the video incident videos can be collected and annotated. Therefore changes to the privacy law should be made as it is currently not allowed to stream and capture incident videos without permission of the building owner.

### 3.2.2 Visual scene type detection

Over the last decade, several literature studies have already shown impressive results regarding the classification of indoor and outdoor scenes [9], and more recent studies have added the ability to predict more detailed attributes, such as the scene type, e.g., kitchen, living room, or bedroom. Zhou et al. [10] constructed a new scene-centric database named Places, containing over 7 million labeled pictures of scenes distributed over 365 classes. Currently, the state-of-the-art systems for scene classification are based on CNNs, mostly trained using a transfer learning approach on the Imagenet models see (as shown in Figure 3.3)[1]. *Transfer learning is a machine learning technique where you 'transfer' the knowledge of the solution from a related task, that has already been illustrated to work, to a new task. Frequently, people pretrain the convolutional network on a large existing dataset (e.g., MNIST, CIFAR, Imagenet or COCO) and then use the network as a fixed feature extractor or use it for initialization before finetuning the final system.*

For the scene classification task the current highest top-5 accuracies (i.e., is the accuracy over all predictions given that the correct class is in the Top-5 highest probabilities of the model) achieved on this Places dataset are 84,91 percent, using the VGG deep CNN [11] and 85,08 percent with the RESNET architecture [7]. Figure 3.4 shows the output of this scene classification approach on indoor scene images.

[1]See Appendix B.3.5

**Figure 3.3** Transfer learning process where you first train the model on a large auxiliary dataset and then finetune the model on a specific target dataset.

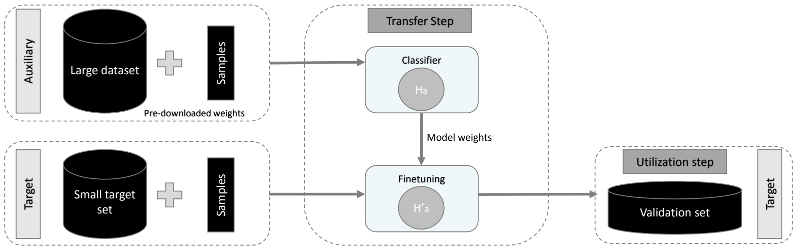



**Figure 3.4** First and second scene classification results (certainty probability between 0 and 1) (source: LSUN-dataset)


### 3.2.3 Visual object localization

In general, object localization consists of two steps: object detection and object recognition (see Figure 3.5). The aim of *object detection* is to find out where an object is located in the image, whereas *object recognition* identifies the semantic class (e.g., couch, chair, table) to which an object belongs. Even though it is a very hot topic in the field of computer vision, there are still some challenges to fully automate the recognition pipeline. The main problem is the large variety of object appearances due to changes in illumination, occlusion, clutter, object pose, viewpoint and non-rigid deformations. A common approach to tackle this issue in the visual domain is to employ CNN architectures [12, 13]. Girshick et al. [14] stated that supervised pre-training on a large auxiliary dataset such as ImageNet [5], followed by domain-specific fine-tuning on a small dataset is an effective paradigm for learning high-capacity CNNs when the available data is scarce. Lu et al. [15] on the other hand showed that transfer learning largely enhances the localization accuracy in the dark environments. Huang et al. [16] gave an overview of the accuracy of the state-of-the-art object detectors (i.e., the single shot detector, the Faster-RCNN architecture and the recurrent fully convolutional network) and their performance.

The following paragraph will give more details on the state-of-art object detectors according to Huang et al. **Faster-RCNN**: Faster-RCNN consists of two modules:

- Region Proposal Network (RPN): gives a set of rectangles based on a deep convolution layer. Basically the RPN slides a small window (3x3) on the feature map, that classifies what is contained inside the window as foreground or background, and also gives the bounding box location of the detected foreground. For every sliding window center it creates a fixed amount of anchor boxes, and classifies these boxes as object or non-object.

- Fast-RCNN Regions Of Interest (ROI) pooling layer: classifies each proposal (from previous step), and refines the proposal location based on the regression loss.

**Single Shot Detector (SSD)**: single shot detector is a first attempt to perform object detection in embedded and regular hardware devices. Instead of having a region proposal framework there is a set of pre-defined boxes to look for objects. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

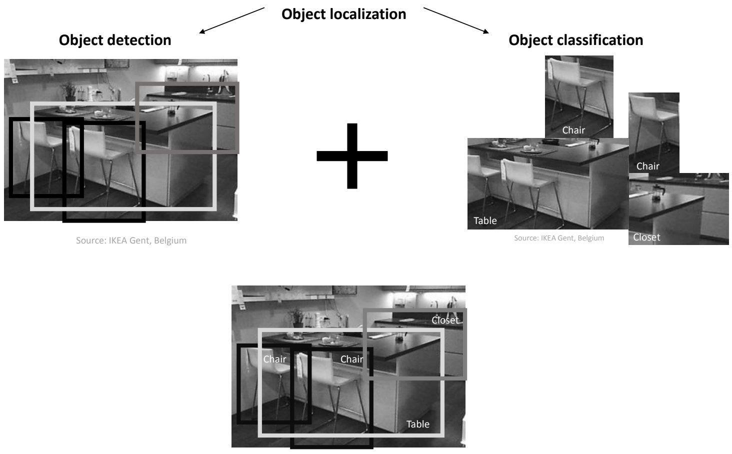
A similar architecture is the YOLO framework [12] where object locations and classes are estimated from one single network. Still the localization errors of the first YOLO version are higher compared to the SSD mechanism. The most recent version YOLOv3 on the other hand replaces the softmax function with independent logistic classifiers to calculate the likeliness of the input in order to belong to a specific label. This makes the object classes not mutually exclusive.

**Region Based Fully-Convolution Neural Network (R-FCN)**: Residual connections allow shortcuts in the model and have allowed to train even deeper neural networks, which have led to even better performance. This has also led to significant simplifications of the Inception blocks. The main difference with the Faster-RCNN architecture is that the region-based detector is fully convolutional with almost all computation shared on the entire image.

**Performance**

In the past (before 2017), most paper tried to increase the accuracy without taking into account the real-time usability and the possibility to integrate the architecture into a low-cost set-up. However, recently there is a huge increase in the usage of deeplearning in real-time applications. Subsection 3.4.3 will go more into detail in the possible real-time optimizations. Only a small set of papers commented on

**Figure 3.5** Object localization consists of two steps: object detection and object recognition. (source original images: IKEA-Ghent)

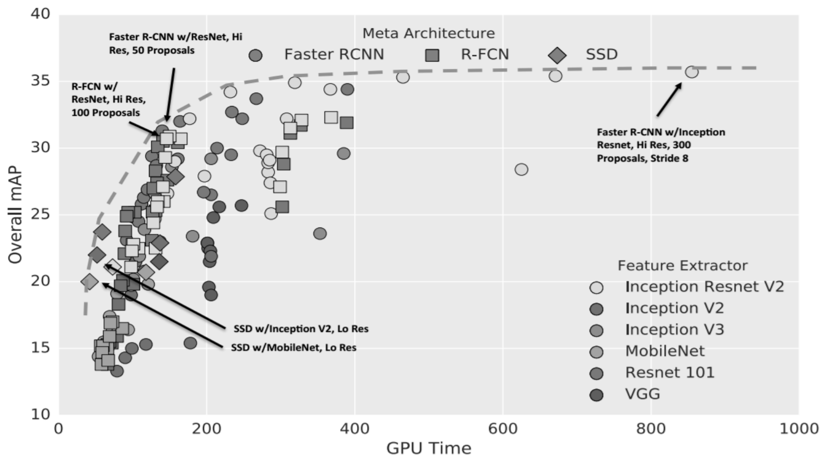



the speed/accuracy trade-off, which depend on different factors (i.e., which feature extractor is used, what is the input image size and how many objects are there on a frame). Based on the results of the paper of Huang et al [16] the following selection criteria are discussed:

- Use the Single Shot Detector (SSD) for best speed performance [17].
- Use the R-FCN and the Faster RCNN for the best balance between speed and accuracy [18].
- Use the Faster RCNN if accuracy is the most important factor. If the number of object proposals is restricted to 50 objects in the Faster RCNN model it is even possible to get similar calculation times as the R-FCN network, without heavily reducing the accuracy. [19].

Figure 3.6 gives a graphical overview of the performance of the Faster RCNN, the R-FCN and the SSD detector framework and their processing time on a GPU device. The detection performance (i.e., 7 frames per second for the Faster RCNN framework on a GPU server) is acceptable for one video, but it is untenable for real deployments at global scale. To put the computational hardware price of a GPU server in context, it would cost over 5 billion USD for hardware alone to analyze all the existing CCTV systems in the UK in real time [20]. To make the detection more feasible we will discuss two optimization methods (i.e., motion detection and object tracking) in Section 3.4.

**Figure 3.6** Mean average precision versus GPU processing time for different networks and object detection models. (source: Huang et al.)




In our framework we want a balance between accuracy and speed. Therefore the loss optimization of the Faster-RCNN network is explained more thoroughly. The Faster-RCNN combines a loss function (see Equation 3.1), the penalty for bad predictions, for two tasks: the localization error and the classification error: $Loss_{=}Loss_{classification} + Loss_{localization}$. The *p* stands for the probability of the box *i* being an object, * is the ground truth label, *t* are the coordinates of the box *i* and *N* are normalization constants.

$$Loss(p_i, t_i) = \frac{1}{N_{cls}} \sum Loss_{cls}(p_i, p_i*) + \frac{\lambda}{N_{box}} \sum p_i * .L_1^{smooth}(t_i - t_i*) \tag{3.1}$$

Although the classification loss is a multi-class loss function (as it needs to predict several class labels), it can be easily transformed into a binary classification loss where you indicate if the class is a particular object class (see equation 3.2) or not. The reason to use the log function is that it heavily penalises classifiers that are confident about an incorrect classification.

$$Loss_{cls}(p_i, p_i*) = -p_i * .logp_i - (1 - p_i*).log(1 - p_i) \tag{3.2}$$

In the visual domain the object detection mechanisms are shown to work, but the study on the application of certain techniques on thermal images is limited. Still recent papers (2018) show the feasibility of object detection in the thermal domain [21, 22]. Finally, infrared object recognition can be vital and beneficial in circumstances with reduced visibility (i.e., due to the smoke in case of a fire).

## 3.3 Thermal object detection

Despite the broad application of handheld thermal imaging cameras in firefighting, its usage is mostly limited to subjective interpretation by the person carrying the device. As remedies to overcome this limitation, object localization and classification mechanisms could assist the fireground understanding and help with the automated localization, characterization and spatio-temporal (spreading) analysis of the fire. An automated understanding of thermal images can transform the conventional knowledge-based firefighting into sensor-driven based firefighting. In this chapter, transfer learning (See Appendix B for more technical details) is applied to multi-labeling convolution neural network architectures for object localization and recognition in monocular visual, infrared images. *Thermal imaging (wavelength around 10 micron) is a particular part of the infrared imaging field (infrared has wavelengths of 0.7 up to 300 micron). Thermal cameras measure the absolute temperature of the object and the cameras are able to work in complete darkness.*

Automatic image recognition is a hot research topic in the field of visual data analysis, but there is less study on the application of such techniques on infrared images. Some examples of applying CNN architectures on infrared images are the work of Wu et al. [23, 24] who focused on the detection of humans in dark environments. Janssens et al. [25] that used infrared images to perform fault detection in bearings and Gundogdu et al. [26] who detected suspicious objects in dark maritime scenes.

Thermal images are useful modalities for investigating low-visibility or dark environments, e.g., during a fire when there is limited visibility due to smoke. In buildings, for example, that contain many different objects made of different materials (e.g., wood, plastic, steel), a thermal image can be used to analyze the amount of heat radiated by each material. For many objects, these radiation profiles are unique, making it possible to recognize objects independent of the luminance. For this reason, the focus of this section is on the recognition of building interior objects by means of handheld thermal images, i.e., a research topic that has limited been addressed so far [27].

### 3.3.1 Thermal and visual aligned dataset creation

The major drawback of a supervised learning algorithm is the need for labeled training data. There are two major options: you can use publically available datasets or you can create your own dataset. The open-source datasets have mostly a higher annotation accuracy, but the annotations are not always detailed or specific enough for a new classification or recognition task. Some well-known public datasets are:

- Common Objects In Context (COCO) [28]: large-scale object detection, segmentation, and captioning dataset.

- ImageNet [5]: single image annotation organized by the semantic hierarchy of WordNet.

- Cityscap dataset [29]: stereo video sequences recorded in street scenes from 50 different cities, with high quality pixel-level annotations.

In order to create your own datasets with highly specific annotations, you can annotate the data yourself, which is a slow process, but the quality is more guaranteed. On the other hand you can use the wisdom of the crowd (i.e., ask several people online, 'in the crowd' to annotate the dataset). Although the quality often suffers, the processing speed increases significantly. Furthermore, there are quality measurements, (such as, multiple person annotation checks) to improve the general quality. There are different online platforms available that offer image annotation tasks online, such as: Crowdflower [30] that allows the upload of video resources, images and text files; Amazon Mechanical Turk[2] and Google image labeler[3].

In order to evaluate the proposed infrared object detection methodology, a dataset is constructed with thermal and visual images [31]. To the best of our knowledge, it is the first indoor scene dataset that combines different spectra of indoor objects. Vidas et al. [32] introduced a related dataset with aligned RGB-D and thermal images from building interiors, but no bounding box annotations were provided, making the dataset useless for object detection verification. Furthermore, the dataset focuses too hard on office scenes making it not useful to annotate them manually.

The images of our dataset are captured with a Flir One[4] thermal camera device. The main advantage of this camera is that thermal and visual images are aligned. This enables us to easily transfer the annotations in the visual domain to the thermal domain without any further calibration, outlining or rectifying. This is a major

[2] https://www.mturk.com/mturk
[3] https://crowdsource.google.com/imagelabeler
[4] http://www.flir.eu/flirone/ios-android/

advantage compared to the previous visual-thermal datasets, in which a checkboard with high contrast in thermal wavelengths was used [33] to calibrate both cameras. It is important to mention that the samples of the current dataset have been collected in standard indoor, non-heated circumstances. As part of future work, we will investigate the impact of the heating during a fire on the object localization performance. In this regard, several real fire footages will be collected and annotated. Furthermore, novel multispectral devices [34] will also be evaluated in the fire investigation context.

The visual domain samples of our dataset were annotated with a bounding box and semantic class label with the well-known LabelMe toolkit [35]. The annotations are then copied to the thermal images. Table 3.1 lists the amount of samples per class. Finally, it is important to remark that the thermal images are taken with a low-cost handheld device and there could be some artifacts in the images due to clutter or motion instability.

**Table 3.1** Enclosure object dataset statistics: number of samples for each object.

| **Object name** | **Number of samples** | **Object name** | **Number of samples** |
|---|---|---|---|
| Closet | 247 | Window | 57 |
| TV or Screen | 29 | Person | 8 |
| Lamp | 227 | Table | 141 |
| Bed | 30 | Chair | 173 |
| Couch | 141 | Potted plant | 56 |

### 3.3.2 Multi-label convolutional network

The neural network that we use is based on the Faster-RCNN architecture introduced in Section 3.2.3. This architecture initiated the automated region proposal technique for object detection which is different compared to conventional methods that utilize selective search [36] or low-level features such as superpixels [37]. First, we adapted the final Fast-CNN layer to the ten classes in our dataset. Then, the weights of the VGG16 (Visual Geometry Group) [38] pre-trained on the COCO-dataset [28] are used to initiate the architecture. In the next step, the learning rate was reduced and early stopping (i.e., a technique to avoid overfitting by limiting the amount of epochs *one epoch is one pass of the full training set to train*). was used to avoid overfitting (i.e., occurs when the model learns the details and the noise in the training data while having a negative impact on the performance of the model on new data) on the new indoor object dataset. Beside the small learning rate, it is also possible to use a decaying learning rate [39] every few epochs (i.e., every couple of epochs reduce the learning rate), but this did not lead to any major changes/improvements. Finally, the joint approximate training (simultaneously optimize the detection and classification task) was used to jointly train the Region Proposal Network (RPN) module [13] and the Fast-RCNN network. The main advantage of this approach is that the network automatically learns the underlying representation of the data without manual intervention.

### 3.3.3 Evaluation and discussion

The fine-tuned network has been evaluated with cross validation on 80% of the dataset while the remaining part (20 %) of the dataset was considered as the test set. The mean Average Precision (mAP) evaluation metric, with values between 0 and 1 (higher is better), is commonly used for the evaluation of object localization. While the average corresponds to the area under the precision-recall curve for a class, the mean of these average individual-class-precision gives the mean Average Precision. To summarize, the mAP score is calculated by taking the mean AP over all IoU thresholds. In addition, a qualitative evaluation of the object detection is performed on the training and test sets. Some examples are shown in Figure 3.7. From these examples we observe that there are more objects detected in the visual images. Also, the certainty of the predicted objects in the visual spectrum is higher compared to the thermal images. Nevertheless, the fact that many objects are correctly detected in the thermal image supports the feasibility of our proposed methodology for thermal object detection.

In order to improve the recognition performance, histogram stretching and gamma correction are used as pre-processing steps to make the images more clear and to increase the contrast. In the case of the thermal images, the mean average precision increased by 22% relatively compared to the original images (without preprocessing). Table 3.2 depicts the mean average precision for the object detection module on the test set. In this table, we also compare the output of the pre-trained weights of the COCO dataset. According to the results, the fine-tuned system significantly improves the traditional COCO system. This confirms the statement of Girshick et al. [14] that *"the fine-tuned network outperformed the randomly initialized pretrained network"*. The results for the thermal images are lower than the visual images, but this was already indicated by the subjective evaluation. Furthermore, there are some objects that radiate the same amount of heat in standard environments which makes the object localization task more difficult. This issue can possibly be solved by using hyperspectral cameras.

**Figure 3.7** Subjective evaluation of the object detection framework in the Visual, Thermal

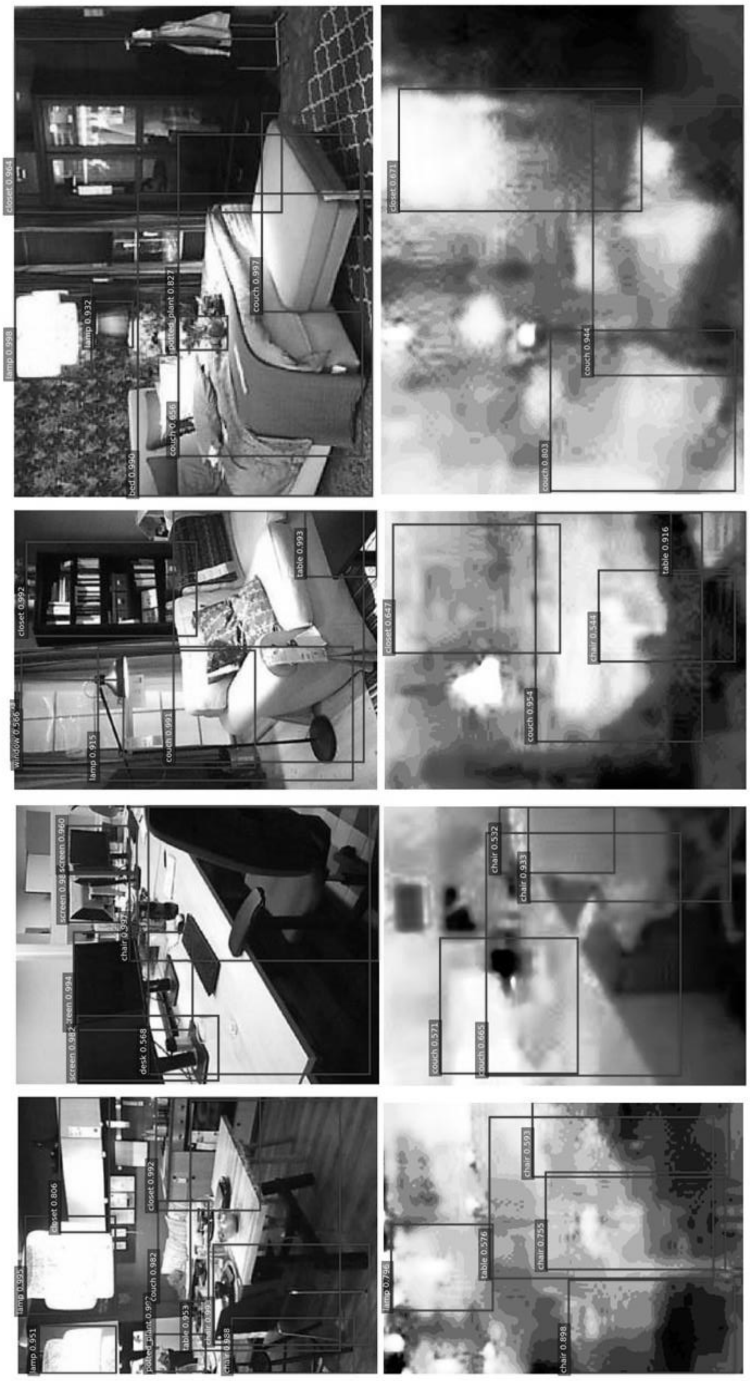


**Table 3.2** Mean average precision (average over multiple Intersection over Union (IoU), where the minimal overlap region is 50%) on the test dataset

| **Object name** | **mAP COCO visual** | **mAP fine-tuned visual** | **mAP thermal** |
|---|---|---|---|
| Bed | 0.171 | 0.418 | 0.143 |
| Chair | 0.102 | 0.348 | 0.027 |
| Closet | 0.006 | 0.237 | 0.024 |
| Couch | 0.330 | 0.381 | 0.108 |
| Lamp | Not in the dataset | 0.171 | 0.059 |
| Person | 0.035 | 0.500 | 0.500 |
| Potted plant | 0.000 | 0.089 | 0.001 |
| Table | 0.000 | 0.281 | 0.023 |
| Tv or screen | 0.250 | 0.129 | 0.018 |
| Window | Not in the dataset | 0.141 | 0.056 |
| Average | 0.112 | 0.269 | 0.096 |

Given that the current dataset is rather small, we propose to further extend the dataset with additional samples from indoor fire scenes. More data samples will most likely increase the final precision of the system. Our main goal, however, was to show the possibilities of our proposed approach.

## 3.4 Real time optimization

The proposed object detection framework is able to process 7 frames per second (from one camera) on a standard GPU device. Optimization mechanisms will be necessary to make a feasible real time system that analyses several static CCTV cameras or a handheld camera without GPU power. The following list gives some suggestions for performance optimization:

- Motion detection (i.e., indicate the frames that have changes in time) via background modeling: the scene and object detection is only run in case there is movement in the field-of-view.
- Object tracking: once the object is detected, you track or follow the object in the camera until the object disappears in the field-of-view.
- Optimized mobile networks, such as mobilenet to reduce the memory consumption of the neural network in an embedded device.
- Contextual and temporal exploitation (see Section 3.5): given that a chair was detected in a frame, the probability to detect a table in a following frame is higher.

**Table 3.3** Performance metrics of evaluated motion algorithms

| **Methods** | **recall** | **Specificity** | **FPR** | **FNR** | **Precision** | **F-measure** |
|---|---|---|---|---|---|---|
| GMM | 0.6226 | 0.9687 | 0.0313 | 0.3775 | 0.4986 | 0.4760 |
| KDE | 0.6866 | 0.9195 | 0.0806 | 0.3134 | 0.3764 | 0.4075 |
| CodeBook | 0.3855 | 0.9888 | 0.0112 | 0.6145 | 0.6119 | 0.4105 |
| AGMM | 0.5603 | 0.9719 | 0.0274 | 0.4397 | 0.6342 | 0.5389 |
| SACON | 0.3822 | 0.9449 | 0.0551 | 0.66178 | 0.4626 | 0.2396 |
| Vibe | 0.4651 | 0.9868 | 0.0131 | 0.5349 | 0.653 | 0.4718 |
| PBAS | 0.5079 | 0.992 | 0.008 | 0.4921 | 0.709 | 0.5505 |

**Table 3.4** Processed frames per second for the evaluated motion algorithms

| Size | GMM | KDE | CodeBook | AGMM | SACON | Vibe | PBAS |
|---|---|---|---|---|---|---|---|
| 320 x 320 | 47.9 | 51.1 | 53.1 | 61.8 | 22.2 | 61.8 | 20.7 |
| 540 x 360 | 21.1 | 35.1 | 19.6 | 53.3 | 7.8 | 52.5 | 14.6 |
| 720 x 480 | 14.7 | 9.5 | 15.2 | 29.2 | 4.2 | 31.5 | 8.1 |

### 3.4.1 Background modeling and movement detection

In order to extract the background, and eventually detect the motion, the video scene should be accurately splitted in foreground (motion) and background (static). We compare different algorithms using a set of standard accuracy metrics and investigate their computational requirements on a publicly available dataset.

The test procedure for the motion estimation is based on the results of the work of Xu et al. [40]. Xu compared eight different algorithms (GMM, KDE, CodeBook, AGMM, SACON, SOBS, Vibe, PBAS) on the Change Detection Benchmark (CDB) [41]. In Table 3.3 and Table 3.4, a general overview of the performances and computational properties of all algorithms is given. They conclude that AGMM, SOBS, Vibe, and PBAS give the most promising results. However, they noticed that sophisticated methods do not always produce more precise results.

In [42] a comprehensive review is given of the performance of all algorithms present in the BGSLibrary [43]. This library provides an easy-to-use C++ framework based on opencv [44] to perform foreground-background separation in videos. The algorithms are evaluated on the Background Models Challenge (BMC) dataset. This dataset is composed of 20 synthetic videos and 9 real videos. Their main conclusion is the high computational cost of their five best methods.

We conducted our own experiments on a subset of the Change Detection Benchmark, namely the subcategories Baseline (containing four videos) and Intermittent Object Motion (containing six videos) to enable a proper evaluation based on a combination of accuracy and speed. We focus on a number of default algorithms present in the well-known library opencv as well as some other more recently developed methods.

### Change detection benchmark

This dataset contains 11 video categories (Baseline, Dynamic Background, Camera Jitter, Intermittent Object Motion, Shadow, Thermal, Bad Weather, Low framerate, Night Videos, PTZ, Turbulence) with 4 to 6 videos sequences in each category. As stated above, we only make use of the category Baseline and Intermittent Object Motion. The motivation behind this is the fact that our video data (indoor video scene data) is mostly similar to these two categories, while the other categories show much less resemblance to it and would possibly confuse our results. Still the other results could be valuable for the evaluation of handheld video footages (i.e., handheld videos can contain camera jitter, low framerate, turbulence and dynamic backgrounds). Finally, the ground truth images contain 5 labels (i.e., 0: static, 50: hard shadow, 85: outside region of interest, 170: unknown motion, 255: motion).

### Methods

We include a number of existing algorithms in our comparative analysis. The extensive description of these methods can be found in their corresponding papers. Furthermore, we also tested the majority of algorithms available in the bgslibrary [45].

*MEDIAN BACKGROUND MODEL [46]* The background is modeled based on the median values of a buffer with a predefined length. The reactivity of the method to changing backgrounds is depending on the length of this buffer. In order to reduce the computational resources the adaptivity of the background model can be lowered by reducing the buffer size.

*ADAPTIVE MEDIAN BACKGROUND MODEL [47]* The background model is a running median of the image sequence produced by the following method: each pixel in the reference image is incremented by one if the corresponding pixel in the current image is greater in value or decreased by one if the current image pixel is less in value.

*MIXTURES OF GAUSSIANS [48]* The background model models each pixel as a mixture of Gaussians and uses an on-line approximation to update the model. In this technique, it is assumed that the intensity values of every pixel in the video can be modeled using a Gaussian mixture model. The pixels that do not match to these distributions are predicted as foreground pixels.

*GMG [49]* This algorithm combines statistical background image estimation and per-pixel Bayesian segmentation. It uses the first few frames for background modelling. It employs a probabilistic foreground segmentation algorithm that identifies possible foreground objects using Bayesian inference. The estimates are adaptive and newer observations are more heavily weighted than old observations to accommodate variable illumination. Several morphological filtering operations like closing and opening are done to remove unwanted noise.

**Evaluation and results**

For each algorithm the average over all videos of the following measures is computed: accuracy, precision, recall, F1 measure and average processing time per frame. This is done by comparing the predicted foreground with the given ground truth for each frame at pixellevel. As can be noted in Table 3.5, the best performing methods on the basis of the F1 score have a low accompanying FPS rate. As mentioned in the literature there is a trade-off between accuracy and speed. MedianBackgroundModel, Vibe and AdaptiveSelectiveBackground are among the best performing algorithms, while also performing at a relatively high FPS rate.

In order to improve the performance of the system and to increase the accuracy of the locations with motion an additional post-processing is performed. This step consists of contour detection, bounding box area calculation and a non-maximum suppression. Contour detection is a technique to find the boundaries (represented as a rectangle) of continuous points in the frame (e.g., points having same color or intensity). Area checking is used to avoid detected blobs or bounding boxes larger or smaller than a preset value. Depending on the camera position this parameter should be evaluated and changed for a particular context (i.e., the detection area will be smaller for person detection in a shopping mall versus person detection in a small hallway). Furthermore, from the scene detection step it is even possible to estimate the maximal or minimal detection area. Currently, the maximum area of a bounding box is set to 25% of the screen area. Non-maximum suppression is a common technique in computer vision to combine/merge highly related detection windows that belong to the same object. High scoring detections are maintained whereas close-by less confident neighbors are removed since they are likely to cover the same object. Refinement and evaluation is needed for each particular parameter of the motion estimation for each specific context. Finally it is important

**Table 3.5** A list of the best-performing algorithms evaluated based on a precision and recall measures, as well as an evaluation on the basis of computational efficiency

| | Accuracy | Precision | Recall | F1 Measure | FPS |
|---|---|---|---|---|---|
| **GMG (opencv) (1)** | 0.93437 | 0.17077 | 0.62089 | 0.26787 | 44.09560 |
| **GMG (opencv) (2)** | 0.96235 | 0.23401 | 0.49451 | 0.31769 | 49.10144 |
| **MOG2 (opencv) (3)** | 0.97968 | 0.41131 | 0.41341 | 0.41236 | 87.52735 |
| **MOG2 (opencv) (4)** | 0.98199 | 0.44121 | 0.42744 | 0.43422 | 116.97275 |
| **DPAdaptiveMedian (5)** | 0.98411 | 0.52842 | 0.38515 | 0.44555 | 158.22785 |
| **DPEigenbackground (6)** | 0.94803 | 0.29977 | 0.90610 | 0.45050 | 27.64875 |
| **MedianBackgroundModel (7)** | 0.97977 | 0.40901 | 0.52516 | 0.45987 | 69.85679 |
| **MedianBackgroundModel (8)** | 0.97964 | 0.40979 | 0.55486 | 0.47142 | 19.05815 |
| **MedianBackgroundModel (9)** | 0.96944 | 0.36907 | 0.78911 | 0.50292 | 97.86651 |
| **ViBe (10)** | 0.98433 | 0.50721 | 0.50491 | 0.50606 | 143.43087 |
| **LBP_MRF (11)** | 0.98144 | 0.43940 | 0.61078 | 0.51111 | 8.14657 |
| **MedianBackgroundModel (12)** | 0.97060 | 0.38243 | 0.78861 | 0.51507 | 27.73771 |
| **DPZivkovicAGMM (13)** | 0.98269 | 0.47578 | 0.56190 | 0.51527 | 83.61204 |
| **AdaptiveSelectiveBackgroundLearning** | 0.97835 | 0.40101 | 0.80839 | 0.53609 | 189.86140 |
| **IndependentMultimodal** | 0.98701 | 0.50056 | 0.57860 | 0.53676 | 27.99630 |
| **LOBSTER (16)** | 0.98692 | 0.53646 | 0.61372 | 0.57250 | 6.33738 |
| **SuBSENSE (17)** | 0.98782 | 0.56930 | 0.63511 | 0.60041 | 3.97388 |
| **PAWCS (18)** | 0.98807 | 0.55798 | 0.75730 | 0.64254 | 1.96115 |

to remark that the usability of the motion detection step to reduce the computational object and scene detection cost depends on the use case. In case the camera is used in a region with high motion, the motion detection is neglected (e.g., during opening hours in shopping mall it is maybe not afforded to run the motion detection step as every region will be classified as motion and every frame will be sent to the object and scene recognition module). On the other hand by continuously running the motion detection algorithm it is possible to have an overview of the regions that have changed over time and only that regions should be investigated and verified with the existing building model (see previous chapter).

### 3.4.2 Object tracking

The object detection should have a decent performance on the whole video (e.g., the couch should be recognized as a couch even if there are people sitting on the couch). To increase the false negatives due to occlusion or low detection probabilities a tracking mechanism is included in the software. When we are tracking an object that was detected in the previous frame, we know a lot about the appearance of the object and we don't have to detect the object again in the next frame. We also know the location in the previous frame and the direction and speed of its motion (e.g. for a person). So in the next frame, we can use all this information to predict the location of the object in the next frame and do a local search around the expected location of the object to accurately locate the object.

It is common for tracking algorithms to accumulate errors and the predicted bounding box often slowly drifts away from the real object it is tracking. Due to the unpredictability and the non-common followed paths in fire behavior content this ratio is currently rather high. To improve the performance of this tracklet *(i.e., the trajectory of the center of the object predictions in previous frames)* it is possible to retrain the Kalman-filter settings according to the specific use case.

A Kalman filter is an optimal estimator and infers parameters of interest from indirect, inaccurate and uncertain observations. It is recursive so that new measurements can be processed as they arrive. Firstly there is some prior knowledge, on which the prediction is based. Secondly, the prediction is updated and refined with new measurements. Thirdly, the output is produced and this result is also used as prior knowledge for the next timestamp. More information and mathematical details can be found in the work of Durrant et al. [50]

Within the opensource software opencv there are 6 tracking mechanisms included: Boosting, MIL, KCF, TLD, MedianFlow an GoTurn. Still none of them is currently able to perform multi object tracking (e.g., tracking multiple persons on a security camera). They all need some kind of initialization and a parameter when the tracking is lost. In order to solve the multi-object tracking problem the Hungarian method of Kuhn et al. [51] is used in our work.

The Hungarian method (or Munkres assignment algorithm) [51] is an optimization algorithm that solves the assignment problem. The assignment problem consist of a number of existing tracklets and a number of new detections. Any detection can belong to any tracklet, incurring some cost that may vary depending on the tracklet-detection assignment (e.g., closeness or smoothness of the tracklet curve). Furthermore, it is required to link all detections by assigning exactly one tracklet to each detection and exactly one detection to each tracklet in such a way that the total cost of the assignment is minimized.

**Simple Online Real-Time (SORT) tracking**

The paper of Bewley et al. [52] reports that the detection quality has a significant impact on tracking performance when comparing different detection mechanisms. Methods which achieve the best accuracy tends to be the slowest. SORT combines the two desirable properties: speed and accuracy without the typical drawbacks. A Kalman filter and a Hungarian [51] method are used to handle the motion prediction and data association components of the tracking problem respectively. When a detection is associated to a target the detected bounding box is used to update the target state where the velocity components are solved optimally via a Kalman filter framework.

The task of object tracking is to create an unique ID when an object enters the image and destroy the ID when the object leaves the image. The new tracker undergoes a probationary period where the target needs to be associated with the detections to accumulate enough evidence to prevent tracking of false positives. Tracks are terminated if they are not used for a specific amount of frames (T-lost). It is better to use a low value of T-lost and create new trackers then having many unnecessary tracks. Furthermore, new tracks are created by any detection with an overlap less than a predefined value. The major change in our framework is the increased value of T-lost. Initially in the paper it was set to 1, but this requires a new identity for objects reappearing. A new identity would also require a new verification of the bounding box (new object detection on the frame each second).

Recently an improved version of the SORT method was proposed by Wojke et al. [53]. By incorporating information through a pre-trained association metric the algorithm is able to track through longer periods of occlusion. This effectively reduces the number of identity switches. Due to the descriptor generation with the residual convolutional neural network the processing speed is reduced to approximately 20 Hz where roughly half of the time spent was on feature generation. More research is necessary to optimize this algorithm for a real-time environment.

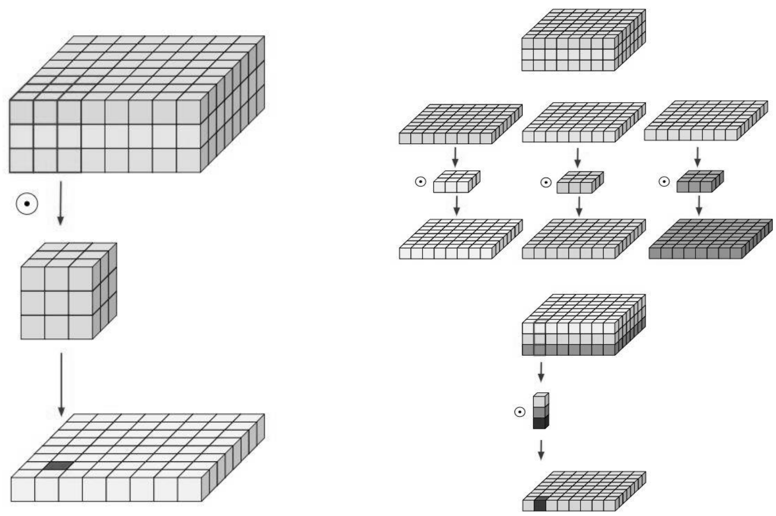
The proposed mechanisms for motion detection and object tracking are useful for our static (i.e., fixed position and field-of-view) CCTV security cameras, but not for handheld cameras due to the high motion. The next subsection will describe a CNN model solution for real-time object detections in embedded or handheld devices (as the tracking and the motion mechanism does not work in handheld cameras due to the global motion). The combination of faster recognition models and motion detection mechanisms (e.g., indicate the rooms with configuration changes) will also be used in static cameras to have a scalable system on all existing security cameras.

### 3.4.3 CNN model optimization

In the beginning of this chapter we described several CNN architectures suitable for classification and object detection tasks. The accuracy of certain algorithms is high, but the networks are not suitable for mobile and embedded vision applications. The MobileNet architecture [54] proposed by Google allows to deploy detection tasks on low power computing resources, for example a smartphone. The main novelty of the MobileNet model is the depth wise separable convolution filter. The standard convolutional filters are split into a depth wise (each RGB channel is kept separate) and a point wise filter (a 1x1 convolution across channels), drastically reducing the parameters and eventually the computation and model size, see Figure 3.8.

The feasibility of the scene detection on the smartphone was tested by retraining the scene detection model of Section 3.2.2. Subsequently, the retrained model was able to predict scenes in real-time on the camera sensor. The scene detection mechanism was retrained on a selection of the Places dataset where the following classes were selected (living room, kitchen, bedroom, dining room, office, bath room and child room). The final accuracy after hyperparameter optimization was 90,6 percent on the trainings dataset, 72,5 percent on the validation and 72,8 percent on the test set. This is lower compared to the initial results of the Resnet and VGG model, but the actual classification speed is much higher. The Mobilenet on an Iphone 7 can process 118 frames per second, which is more than 30FPS similar to a standard video fragment, whereas the VGG network only processed 3 up to 4 images per second (for an input image of 224x224).

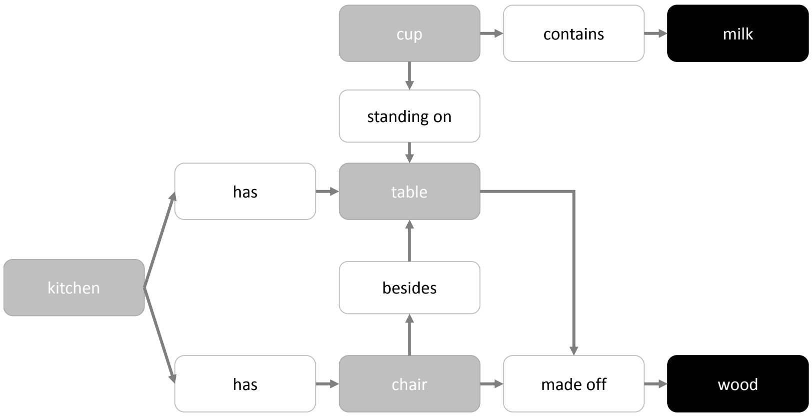
**Figure 3.8** Standard convolution operation on a 3-channel image (left) and depth wise separable convolution operation on a 3-channel image (right), (source: eli.thegreenplace.net)



## 3.5 Probabilistic scene understanding

Besides the performance optimization, which was discussed in the previous section, it is important to improve the accuracy of the scene and object detections (e.g., a sleeping room was quite often confused with a living room). This is firstly possible by exploiting the context (i.e., the probability to predict a bed, given that the room is a sleeping room is much higher compared to a kitchen and vice versa). Secondly, in case the amount of objects in a particular scene, in the field-of-view of the camera, are too small (e.g., a close-up shot), the accuracy of the scene predictions results will be lower and they are then neglected. Finally, it is possible to create a scene graph (see Figure 3.9) to describe the correlations between the objects (e.g., *a cup is standing on the table*). Johnson et al. defined a scene graph [55], as a structured representation of an image, where nodes in a scene graph correspond to object bounding boxes *(e.g., table, cup)* with their object categories and edges correspond to their pairwise relationships between objects *(e.g., is standing on)*.

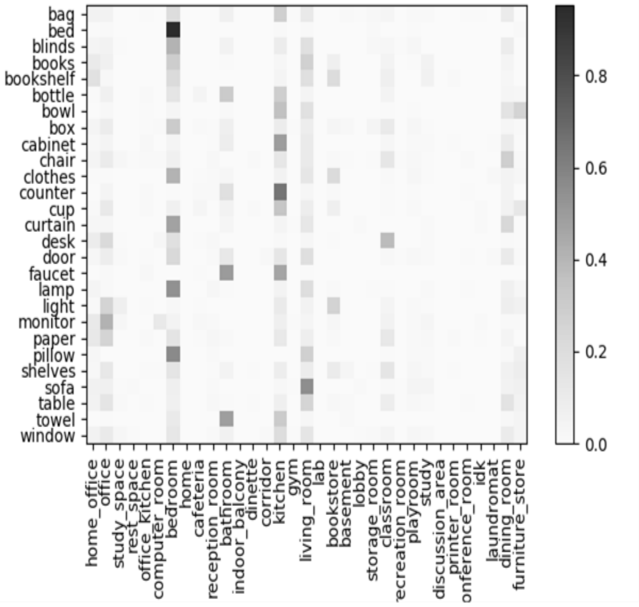
**Figure 3.9** Scene graph visualization: ”The chair is besides the table in the kitchen and the cup that contains milk is standing on the table.”




### 3.5.1 Contextual scene exploitation

Intuitively, individual predictions of objects and relationships can benefit from their surrounding context. In our work we optimized the scene classification results by exploiting the object detections. First we learned the object co-occurrence from the ObjectNet3D dataset [56]. Some preprocessing was necessary to remove the objects that appear only once in the dataset. Furthermore the classes floor, ceiling and wall were removed from the object dataset as they are not descriptive for a specific room type. Secondly, the object-scene occurrence was derived from the SUN RGB-D dataset [57] (annotated with the scene tag and the objects). The visualization of the major classes of the object-co-occurrence is shown in Figure 3.10 and the scene and object occurrence is given in Figure 3.11.

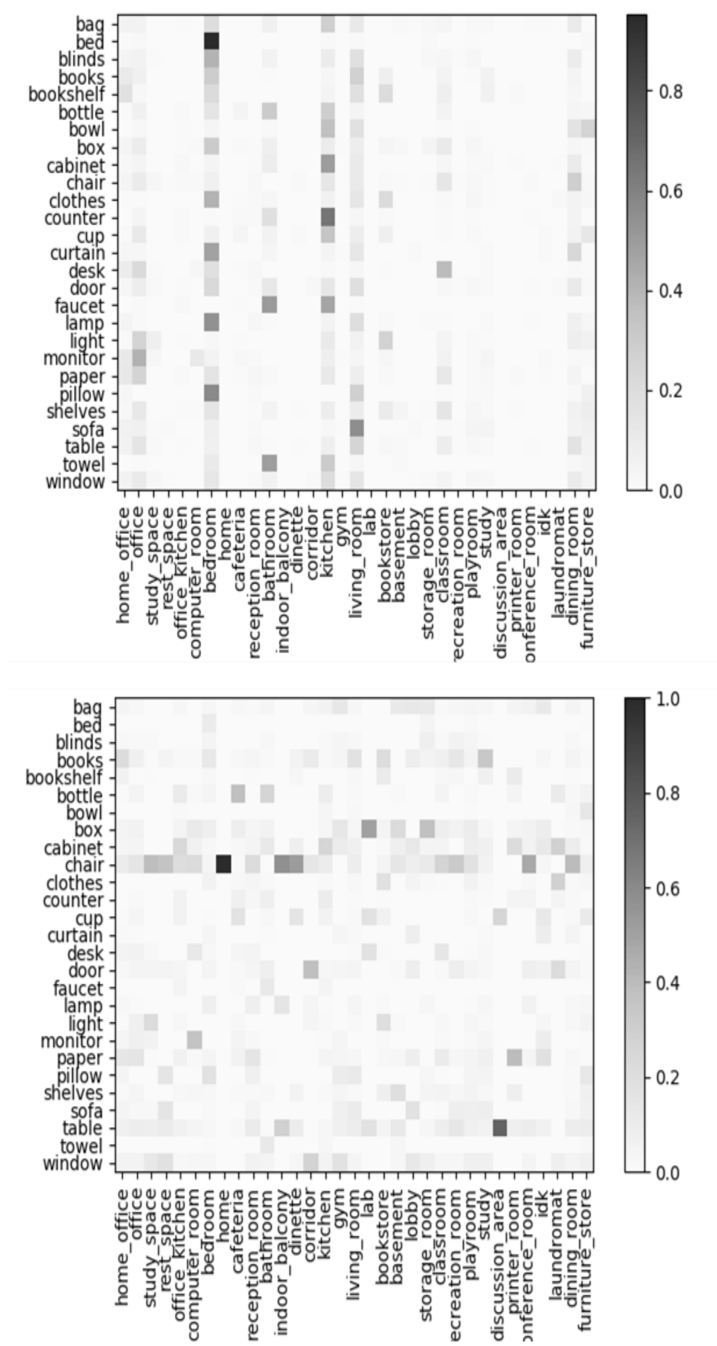
**Figure 3.10** Object co-occurrence according to the SUN RGB-D dataset




The usability of the contextual process is twofold. Firstly the results of the scene detection are improved (in case there is high uncertainty) by exploiting the probability of the detected objects (scene occurrence). The 5 major scene detection classes are refined by taking the normalized score of the product of the particular scene and the sum of the weighted object scores.

Secondly, the object-co-occurrence will assist the thermal object detection by indicating the objects that are related to the detected classes. For example if a chair is detected, there is a 60 percent chance that there will be a table in the same scene. This can then be incorporated in the fire load calculation. Although there are still errors possible, these mechanisms allows us to estimate more in detail the fire complexity. Finally, the co-occurrence is currently derived from several object and scene datasets, but the weights could be learned from labeled data. The loss function would be the correctness of the scene label given a particular set of objects.

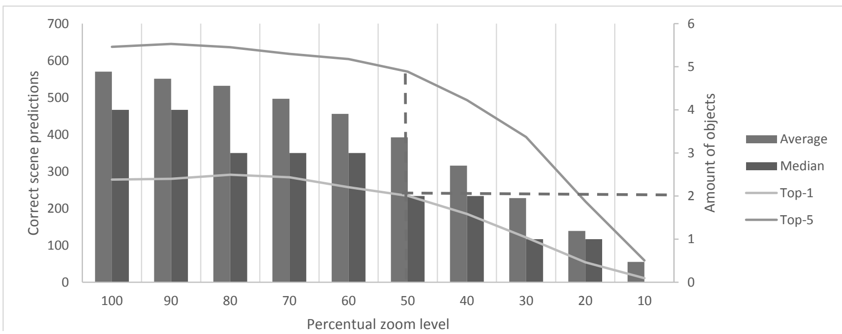
**Figure 3.11** Scene occurrence, normalized by the scene tag (top), normalized by the object tag (bottom).



### 3.5.2 Scene location validation

The place recognition is working perfectly for frames with a lot of objects in the camera. Still, for close-up shots the place detection algorithm fails to give decent prediction results. It is hard to quantify the particular amount of objects necessary to have a specific scene classification result. An evaluation was done on the NYU scene [58] by investigating the impact of the zoom level of several ground truth labeled scenes and the amount of particular objects that were detected with at least 50 percent certainty. As can be seen in Figure 3.12 there is a strong correlation between the amount of detected objects and the certainty of the scene labels. If the amount of detected objects is too low (i.e., less then 2 objects with 25 percentage certainty detected), the scene detection result is neglected and the following frames (that are zoomed out) should be used for scene detection. Instead of the camera sensor, other information sources (i.e., ventilation conditions [59] or lightning set-up [60]) could be used to determine or verify the room type.

**Figure 3.12** Amount of detected objects (with 25 percentage certainty) in function of the scene type accuracy on the LSUN dataset.



## 3.6 Object metadata for fire understanding

The last section in this chapter will discuss the required object metadata for fire understanding. Besides the object and the scene labels we need (at least) information of the 3D-cuboid of the object (i.e., the orientation, the volume and the position) and the material where the object is made of. In this section we propose a methodology to incorporate this information in our framework. The implementation of this final step, however, is outside the scope of this thesis. Finally, some details are given on person detection as cameras are highly valuable to give information about the presence of persons in the building.

### 3.6.1 3D object recognition

3D object detection is fundamental for detailed scene understanding (in case the object dimensions and details are not available in the building model). Over the past years many steps are taken in the 3D machine learning field which is an interdisciplinary field that fuses computer vision, computer graphics and machine learning. Initially, the focus was mainly on multi-sensor approaches (more specifically on stereo-vision), but since last year more and more data driven approaches are proposed where the 3D boxes or the pose-estimation are derived from large datasets (for example the LSUN July 2018 challenge[5], were 13000 indoor 3D scenes were annotated). Mousavian et al. and Izadinia et al. [61, 62] proposed a technique for estimating the 3D object detection and pose from a single image. Luo et al. [63] combined the depth and the RGB image to make decent 3D detections. In general the problem can be described as a 6 degrees of freedom problem (pose, position and dimensions of the object). Two main optimization solutions are discussed in literature. Firstly, the detection results of Section 3.2.3 can be passed as input to a pose estimation network. Secondly, the network can jointly perform viewpoint estimation and 2D detection with a combined loss function. Finally, we suggest to incorporate the work of Mousavian et al. for 3D object recognition in our global scene understanding framework as it is possible to estimate the pose and the dimensions from a single image (e.g., a standard security camera). It is important to remark that the 3D recognition module should only be run (due the computational cost) in case the motion and 2D object detection module recognize changes in the scene configuration.

[5] `http://rgbd.cs.princeton.edu/challenge.html`

### 3.6.2 Material type recognition

Besides the object type and the 3D cuboid, it is necessary to have an indication of the material type before you can incorporate it in a fireforecasting system. A chair made of plastics will have another fire curve than a hard wooden chair. In literature several datasets are available: the Flickr Material Database [64] consists of color photographs of surfaces belonging to one of ten common material categories: fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood. The FOD dataset [65] is a more challenging set with plastic, metal and stone elements in challenging and deformed conditions. Wieschollek et al. [66], Kalliatakis et al. [67], and Bell et al. [68] indicated that deep neural networks are suitable for the visual material type classification task with the available amount of labeled data. Furthermore, Cho et al. [27] created recently (2018) an automatic material recognition task with good performances (above 89 % mean accuracy from 15 indoor and 17 outdoor materials). The images were retrieved using a low-cost mobile thermal camera integrated into a smartphone to capture thermal textures and the images were collected within various temperatures. This allows to detect the material type even in darkness or in a smoke, non-visible surrounding. With the increased availability of thermal cameras there is a huge opportunity for increased data collection in various challenging conditions. Finally the material recognition step can be used to update or enrich the BIM model with more detailed material information (e.g., the chair is made for 80 % of plywood and 20 percent of aluminum).

### 3.6.3 Person detection

Person (i.e., victim or evacuated people) detection are popular topics in computer vision and different visual and multispectral benchmarks have been proposed in literature [69]. The object detection mechanisms proposed in this chapter are, besides detecting furniture, able to perform person detection. Another alternative is to use engineered features or to train a simple CNN network that acts as a binary classifier, more specific, is the bounding box a person yes or no (the box is the region where there is motion detected in the scene).

Following paragraph will shortly evaluate the performance and the accuracy of the feature engineered and feature learned based classifiers. The classifier was trained on the INRIA training set [70], validated on the INRIA training set and tested on the TUD-Motionpairs Pedestrian dataset [71]. Four models/techniques were evaluated. Firstly the standard HOG and SVM combination. Secondly the HOG feature extractor with an additional neural network consisting of two layers for classification. Thirdly, the convolutional neural network consisting of 6 convolutional layers and two fully connected layers.

**Table 3.6** Person detection results, evaluated on the INRIA and TUD-Motionpairs dataset.

| | *Training error* | *Validation error* | *Test error* | *FPS* |
|---|---|---|---|---|
| **HOG + SVM** | 0.0033 | 0.0648 | 0.0902 | 107 |
| **HOG + 2-layer NN** | 0.0213 | 0.0486 | 0.0928 | 314 |
| **CNN** | 0.0147 | 0.0299 | 0.0817 | 280 |
| **Pre-trained CNN** | 0.0055 | 0.0027 | 0.0352 | 95 |

Fourthly, the fine-tuned VGG- network on the ImageNet dataset. From Table 3.6 it can be clearly seen that the accuracy of the fine-tuned CNN network is the highest. Still by comparing the accuracy gain versus the processing rate the CNN layer is chosen in the architecture.

Video based solutions are usually applied in public spaces but are not acceptable in private spaces such as Smart Homes for obvious privacy reasons. Furthermore, they are not robust to occlusion and do not offer any way of recovering from undetected events. An alternative for person detection and privacy ensuring is to use radar technology as proposed by Zhao et al. [72] or to use reasoning on other home sensors: Renoux et al [73] and Marroquin et al. [74].

## 3.7 Conclusions and future work

In this chapter, we proposed a framework for automated fireground understanding from images. Furthermore, we investigated the application of multi-labeling CNN's on object and scene detection in visual and infrared images. Moreover the real-time optimization of the scene understanding framework was improved by using motion detection, object tracking and model optimization. The usability of the proposed framework is twofold. Firstly, the scene and object understanding could facilitate firefighting robots as they need to understand the surrounding before taking actions. Secondly, indoor scene understanding can optimize or confirm the current inside information available in the building information model. Future work will focus on the optimization of the object detection, with more recent techniques and scene detection in challenging conditions. On the one hand an automated collection of infrared images from real fires could facilitate the data collection process. On the other hand Yun et al. [75] showed that it is possible to generate thermal information from a visual body camera with conditional generative adversarial networks. Subsequently, 3D object and scene recognition should be further investigated with adapted CNN networks. Finally, future work will need to incorporate the scene and object detections in building information models and fire forecasting frameworks.

# References


[1] Dana H Ballard, Christopher M Brown, and Jerome A Feldman. *An approach to knowledge-directed image analysis*. Technical report, ROCHESTER UNIV NY DEPT OF COMPUTER SCIENCE, 1977.

[2] Aravindh Mahendran and Andrea Vedaldi. *Understanding Deep Image Representations by Inverting Them*. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[3] Karen Simonyan and Andrew Zisserman. *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.

[4] Justin Johnson. *Benchmarks for popular cnn models*, 2016.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks*. In Advances in neural information processing systems, pages 1097–1105, 2012.

[6] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. *Overfeat: Integrated recognition, localization and detection using convolutional networks*. arXiv preprint arXiv:1312.6229, 2013.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[8] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. *Inception-v4, inception-resnet and the impact of residual connections on learning.* In AAAI, volume 4, page 12, 2017.

[9] Jiebo Luo and Andreas Savakis. *Indoor vs outdoor classification of consumer photographs using low-level and semantic features*. In International Conference on Image Processing (ICIP), volume 2, pages 745–748, 2001.

[10] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. *Learning deep features for scene recognition using places database*. In Advances in Neural Information Processing Systems (NIPS), pages 487–495, 2014.

[11] Limin Wang, Sheng Guo, Weilin Huang, and Yu Qiao. *Places205-VGGNet models for scene recognition*. arXiv:1508.01667, 2015.

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. *You only look once: Unified, real-time object detection*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 779–788, 2016.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. *Faster r-cnn: Towards real-time object detection with region proposal networks*. In Advances in neural information processing systems, pages 91–99, 2015.

[14] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[15] Guoyu Lu, Yan Yan, Li Ren, Philip Saponaro, Nicu Sebe, and Chandra Kambhamettu. *Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging*. Neurocomputing, 173:83–92, 2016.

[16] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. *Speed/accuracy trade-offs for modern convolutional object detectors*. In IEEE CVPR, volume 4, 2017.

[17] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. *SSD: Single shot multibox detector*. In European Conference on Computer Vision, pages 21–37. Springer, 2016.

[18] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. *R-fcn: Object detection via region-based fully convolutional networks*. In Advances in neural information processing systems, pages 379–387, 2016.

[19] Ross Girshick. *Fast r-cnn*. In Proceedings of the IEEE International Conference on Computer Vision, pages 1440–1448, 2015.

[20] Sulabh Bhattarai and Yong Wang. *End-to-End Trust and Security for Internet of Things Applications*. Computer, 51(4):20–27, 2018.

[21] Dongdong Zeng and Ming Zhu. *Multiscale Fully Convolutional Network for Foreground Object Detection in Infrared Videos*. IEEE Geoscience and Remote Sensing Letters, 15(4):617–621, 2018.

[22] Christian Herrmann, Miriam Ruf, and Jürgen Beyerer. *CNN-based thermal infrared person detection by domain adaptation*. In Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, volume 10643, page 1064308. International Society for Optics and Photonics, 2018.

[23] Zheng Wu, Nathan Fuller, Diane Theriault, and Margrit Betke. *A Thermal Infrared Video Benchmark for Visual Analysis*. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2014.

[24] Amanda Berg, Jörgen Ahlberg, and Michael Felsberg. *A thermal object tracking benchmark*. In Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on, pages 1–6. IEEE, 2015.

[25] Olivier Janssens. *Data-driven performance monitoring, fault detection and dynamic dashboards for offshore wind farms*. PhD thesis, 2017.

[26] Erhan Gundogdu, Aykut Koç, and A Aydın Alatan. *Object classification in infrared images using deep representations*. In Image Processing (ICIP), 2016 IEEE International Conference on, pages 1066–1070. IEEE, 2016.

[27] Youngjun Cho, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Simon J Julier. *Deep Thermal Imaging: Proximate Material Type Recognition in the Wild through Deep Learning of Spatial Surface Temperature Patterns*. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 2. ACM, 2018.

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. *Microsoft coco: Common objects in context*. In European Conference on Computer Vision, pages 740–755. Springer, 2014.

[29] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. *The cityscapes dataset for semantic urban scene understanding*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3213–3223, 2016.

[30] Panagiotis G Ipeirotis. *Analyzing the amazon mechanical turk marketplace*. XRDS: Crossroads, The ACM Magazine for Students, 17(2):16–21, 2010.

[31] Florian Vandecasteele, Bart Merci, Azarakhsh Jalalvand, and Steven Verstockt. *Object localization in handheld thermal images for fireground understanding*. In Thermosense: Thermal Infrared Applications XXXIX, volume 10214, page 1021405. International Society for Optics and Photonics, 2017.

[32] Stephen Vidas, Peyman Moghadam, and Michael Bosse. *3D thermal mapping of building interiors using an RGB-D and thermal camera*. In Robotics and Automation (ICRA), 2013 IEEE International Conference on, pages 2311–2318. IEEE, 2013.

[33] Philip Engström, Håkan Larsson, and Joakim Rydell. *Geometric calibration of thermal cameras*. In SPIE Security+ Defence, pages 88970C–88970C. International Society for Optics and Photonics, 2013.

[34] Andy Lambrechts, Pilar Gonzalez, Bert Geelen, Philippe Soussan, Klaas Tack, and Murali Jayapala. *A CMOS-compatible, integrated approach to hyper-and multispectral imaging*. In Electron Devices Meeting (IEDM), 2014 IEEE International, pages 10–5. IEEE, 2014.

[35] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. *LabelMe: a database and web-based tool for image annotation*. International journal of computer vision, 77:157–173, 2008.

[36] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. *Selective search for object recognition*. International journal of computer vision, 104:154–171, 2013.

[37] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. *SLIC superpixels compared to state-of-the-art superpixel methods*. IEEE transactions on pattern analysis and machine intelligence, 34:2274–2282, 2012.

[38] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. *Imagenet large scale visual recognition challenge*. International Journal of Computer Vision, 115:211–252, 2015.

[39] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. *How transferable are features in deep neural networks?* In Advances in neural information processing systems, pages 3320–3328, 2014.

[40] Yong Xu, Jixiang Dong, Bob Zhang, and Daoyun Xu. *Background modeling methods in video analysis: A review and comparative evaluation*. CAAI Transactions on Intelligence Technology, 1(1):43–60, 2016.

[41] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. *CDnet 2014: An expanded change detection benchmark dataset*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 387–394, 2014.

[42] Andrews Sobral and Antoine Vacavant. *A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos*. Computer Vision and Image Understanding, 122:4–21, 2014.

[43] A Sobral. *An openCV C++ background subtraction library*. In IX Workshop de Visao Computacional (WVC). Library avalaible at: http://code. google. com/p/bgslibrary/. Accessed, 2013.

[44] Gary Bradski and Adrian Kaehler. *OpenCV*. Dr. Dobbs journal of software tools, 3, 2000.

[45] Andrews Sobral. *BGSLibrary: An opencv c++ background subtraction library*. In IX Workshop de Visao Computacional, volume 2, page 7, 2013.

[46] Daniel Gutchess, M Trajkovics, Eric Cohen-Solal, Damian Lyons, and Anil K Jain. *A background model initialization algorithm for video surveillance*. In Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, volume 1, pages 733–740. IEEE, 2001.

[47] Ahmed Elgammal, David Harwood, and Larry Davis. *Non-parametric model for background subtraction*. In European conference on computer vision, pages 751–767. Springer, 2000.

[48] Anurag Mittal and Nikos Paragios. *Motion-based background subtraction using adaptive kernel density estimation*. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–II. Ieee, 2004.

[49] Andrew B Godbehere, Akihiro Matsukawa, and Ken Goldberg. *Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation*. In American Control Conference (ACC), 2012, pages 4305–4312. IEEE, 2012.

[50] Hugh Durrant-Whyte et al. *Introduction to estimation and the Kalman filter*. Australian Centre for Field Robotics, 28(3):65–94, 2001.

[51] Harold W Kuhn. *The Hungarian method for the assignment problem*. Naval research logistics quarterly, 2(1-2):83–97, 1955.

[52] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. *Simple online and realtime tracking*. In Image Processing (ICIP), 2016 IEEE International Conference on, pages 3464–3468. IEEE, 2016.

[53] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. *Simple online and realtime tracking with a deep association metric*. In Image Processing (ICIP), 2017 IEEE International Conference on, pages 3645–3649. IEEE, 2017.

[54] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861, 2017.

[55] Justin Johnson, Ranjay Krishna, Michael Stark, Li-Jia Li, David Shamma, Michael Bernstein, and Li Fei-Fei. *Image retrieval using scene graphs*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3668–3678, 2015.

[56] Yu Xiang, Wonhui Kim, Wei Chen, Jingwei Ji, Christopher Choy, Hao Su, Roozbeh Mottaghi, Leonidas Guibas, and Silvio Savarese. *Objectnet3d: A large scale database for 3d object recognition*. In European Conference on Computer Vision, pages 160–176. Springer, 2016.

[57] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. *Sun rgb-d: A rgb-d scene understanding benchmark suite*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 567–576, 2015.

[58] Nathan Silberman and Rob Fergus. *Indoor scene segmentation using a structured light sensor*. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pages 601–608. IEEE, 2011.

[59] KWD Cheong, E Djunaedy, KW Tham, SC Sekhar, NH Wong, and MB Ullah. *Influence of furniture layout and ventilation design on air quality and thermal comfort*. 2002.

[60] Seyed Mohammad Mousavi, Tareef Hayat Khan, and Yaik-Wah Lim. *Impact of Furniture Layout on Indoor Daylighting Performance in Existing Residential Buildings in Malaysia*. Journal of Daylighting, 5(1):1–13, 2018.

[61] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Košecká. *3d bounding box estimation using deep learning and geometry*. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 5632–5640. IEEE, 2017.

[62] Hamid Izadinia, Qi Shan, and Steven M Seitz. *Im2cad*. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 2422–2431. IEEE, 2017.

[63] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J Davison. *A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM*. In Robotics and automation (ICRA), 2014 IEEE international conference on, pages 1524–1531. IEEE, 2014.

[64] Lavanya Sharan, Ruth Rosenholtz, and Edward Adelson. *Material perception: What can you see in a brief glance?* Journal of Vision, 9(8):784–784, 2009.

[65] Haoyu Xu, Zhenqi Han, Songlin Feng, Han Zhou, and Yuchun Fang. *Foreign object debris material recognition based on convolutional neural networks*. EURASIP Journal on Image and Video Processing, 2018(1):21, 2018.

[66] Patrick Wieschollek and Hendrik Lensch. *Transfer Learning for Material Classification using Convolutional Networks*. arXiv preprint arXiv:1609.06188, 2016.

[67] Grigorios Kalliatakis, Georgios Stamatiadis, Shoaib Ehsan, Ales Leonardis, Juergen Gall, Anca Sticlaru, and Klaus D McDonald-Maier. *Evaluating deep convolutional neural networks for material classification*. arXiv preprint arXiv:1703.04101, 2017.

[68] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. *Material recognition in the wild with the materials in context database*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3479–3487, 2015.

[69] Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, Yi Yang, Qi Tian, et al. *Person Re-identification in the Wild.*

[70] Navneet Dalal and Bill Triggs. *Histograms of oriented gradients for human detection*. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893. IEEE, 2005.

[71] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. *Pedestrian detection: An evaluation of the state of the art*. IEEE transactions on pattern analysis and machine intelligence, 34(4):743–761, 2012.

[72] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. *Through-wall human pose estimation using radio signals*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7356–7365, 2018.

[73] Jennifer Renoux, Marjan Alizeraie, Lars Karlsson, Uwe Köckemann, Federico Pecora, and Amy Loutfi. *Context recognition in multiple occupants situations: Detecting the number of agents in a smart home environment with simple sensors*. In Workshop on Knowledge-Based Techniques for Problem Solving and Reasoning (KnowProS17), 2017.

[74] Roberto Marroquin, Julien Dubois, and Christophe Nicolle. *WiseNET: smart camera network combined with ontological reasoning for smart building management*. In 5th Workshop on Architecture of Smart Cameras (WASC), 2016.

[75] Kyongsik Yun, Jessi Bustos, and Thomas Lu. *Predicting Rapid Fire Growth (Flashover) Using Conditional Generative Adversarial Networks*. Electronic Imaging, 2018(9):1–4, 2018.

# 4

# Spatio-temporal fire characteristics

*This chapter presents the general architecture of a multi-sensor geographic information system that allows the effective use of sensor data and geographic information in fire incident management and fire development analysis. The proposed platform allows the generation of real-time heatmaps that show the space-time distribution of fire/smoke risk levels across an area of concern based on multi-view sensing. Such levels can assist the decision makers in taking actions and aim at facilitating quick fire emergency response. Results of several real fire experiments in a large-scale road tunnel and in a multi-compartment set-up show the feasibility of the approach. Subsequently in this chapter, more details are given on the smoke visibility estimation algorithm and the flame height derivation from video footages. Furthermore, by temporal analysis of the within- and between-variance of the sensor estimations, an indication on the accuracy/certainty of the measurements at each moment in time can be given. In this way, problems that arise in one sensor can be compensated by the other surrounding sensors. Finally, in this chapter a discussion on the firemap generation is given. User experience studies have shown that the platform improves the current visualization of fire characteristics.*

## 4.1 Introduction

The real utilization of spatial-information, such as road/building maps and real-time traffic data, and its combination with geotagged fire incident data is still limited in the analysis of fire emergency situations [1]. Geographic reasoning on fire events from heterogeneous multi-sensor observations, i.e., the main focus of this chapter, will help the fire crew in their decision-making process [2, 3]. Our fast on-site collaborative data collection and dynamic incident map creation on which space-time visual analysis can be performed, will facilitate future fire operations.

The location, the size and the thickness of smoke can change the strategic and tactical plan at the fire scene as indicated in Chapter 1. As such, reading smoke is essential for early warning and prediction of the fire behavior [4–6]. By observing the spreading characteristics and the thickness of smoke, firefighters can have a better understanding of the conditions that they will face. The proposed fire geographic information system, 'fireGIS' facilitates the smoke reading and automatically measures the fire and smoke characteristics and visualizes them on a spatio-temporal map of the environment. One use case of the spatio-temporal map is the evaluation of real fire experiments, e.g., to compare the impact of sprinklers and smoke evacuation systems.

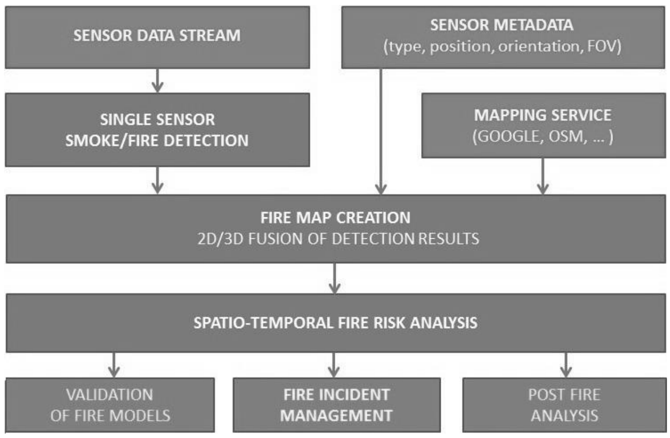
FireGIS builds further on the multi-modal/multi-sensor fire detection work that has been performed during the past years [7–9]. The main focus of preceding approaches was the fast detection and localization of the fire and smoke sources. This work extends previous work with the spatio-temporal mapping of the sensor or video data into real-time heatmaps that show the space-time distribution of fire risk levels. There are three major steps involved in the fireGIS process:

- The collection of low-cost (i.e., computationally efficient) multi-sensor data for the fire risk assessment,
- The fire maps creation,
- The spatio-temporal fire risk analysis.

Section 4.2 will discuss each of these steps in more detail and illustrate their application by means of several real fire experiments that are described in Section 4.3. In these experiments, different cameras were used to monitor the visibility-based smoke features. The visibility measurement is explained in more detail in Section 4.4. Temporal analysis of the within- and between-variance of the sensors' visibility estimations are used to give an indication on the accuracy/certainty of the measurements at each moment in time. In this way, problems that arise in one sensor can be compensated by the other surrounding sensors. This is further explained in Section 4.5. Subsequently, in Section 4.6 more details will be given on the video flame analysis. Furthermore, Section 4.7 presents the fire map generation and Section 4.8 discusses the evaluation and verification of the platform. This is done by a subjective analysis of the visibility level in the videos and an objective comparison to the temperature profiles of a set of thermocouple trees.

## 4.2 General fireGIS architecture

**Figure 4.1** General fireGIS architecture for spatio-temporal fire risk analysis.



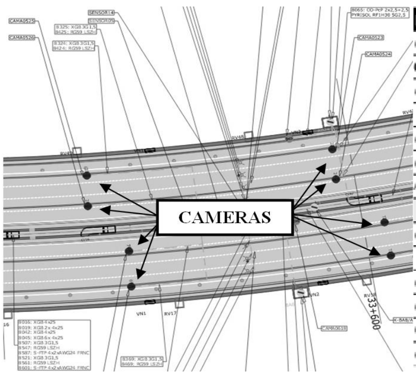

The general architecture of the fireGIS platform is shown in Figure 4.1. In order to start up the fireGIS analysis, the platform needs to get meta data input about the sensors and the environment that needs to be monitored. For each of the available sensors, a link to the sensor data stream and the location information, i.e., position, orientation and field of view (FOV), needs to be registered in the fireGIS platform. In our tunnel experiments, for example, this information was provided by the Agency for Roads and Traffic (AWV) and the Flemish Tunnel and Control Center (VTC). In Figure 4.2, an overview is given about the data as provided by both agencies.

It is important to remark that, in its current form, the data cannot be imported directly into the fireGIS architecture and some pre-processing is necessary. In the future, with the increased use of sensor standardization mechanisms, for example by following the SNN ontology [10] for sensor and observation properties, the data will be incorporated in a more efficient way. Furthermore, the data should be available in the BIM file of the particular building. Finally, the user needs to choose on which GIS or mapping service, e.g., Google Maps or OpenStreetMap (OSM), the spatio-temporal fireGIS detection results, i.e., the output of our platform, need to be shown. It is also possible to map the information onto an existing floor plan of the environment, as shown in Figure 4.3.

In the next step, i.e., after all meta data information is provided, the low-cost smoke analyzing algorithm will start inspecting the visual data streams. This process is further described in Section 4.4. In this thesis, we mainly discuss the use of video data but the generic character of the framework also allows other sensor types to be included. In Section 4.7, for example, we describe the incorporation of temperature data/profiles derived from a set of thermocouple trees. Subsequently, the sensor detection results are projected to a 2D or 3D map of the environment using the location information of the sensors, as shown in Figure 4.2. In order to give an indication of the fire risk, different color codes ranging from green to red are used, corresponding to the detected smoke/visibility at each monitored point/region. For the presented fire experiments, mapping is done to a 2D representation of the environment.

Finally, by analyzing the generated fire risk maps over time, a spatio-temporal analysis can be performed on the fire spreading. This can be very useful real-time information for fire incident management, but can also be used for post fire analysis. It is important to remark that it is not the intention of this Chapter to propose a new fire or smoke detection algorithm. The fireGIS platform is more like a system methodology in which the fire/smoke detection mechanism is a black box independent of the used sensor type, i.e., the detection itself can easily be replaced and extended with other state-of-the-art detectors available in literature [11–13].

**Figure 4.2** Sensor and environment input provided by the Agency for Roads and Traffic (AWV) and the Flemish Tunnel and Control Center (VTC) - Road map with sensor locations (left) and links to sensor data streams and additional positioning/orientation information (right).



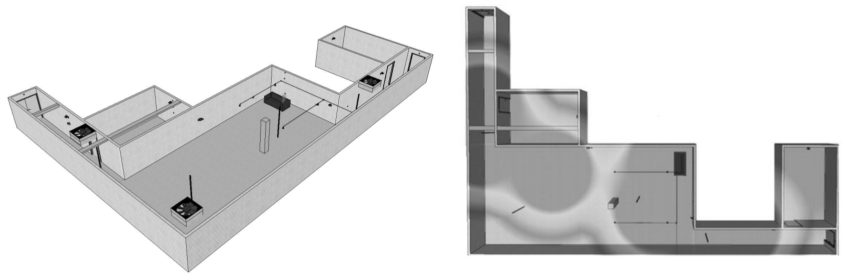

| Filename | Link | Orientation | Positio |
|---|---|---|---|
| CAMA0592.mov | link | SO | 4 |
| CAMA0624p.mov | link | PTZ | 3.5 |
| CAMA0590.mov | link | SA | 3 |
| | | | 2.5 |
| CAMA0588.mov | link | SA | 2 |
| CAMA0586.mov | link | SA | 1 |
| CAMA0622p.mov | link | PTZ | 0.5 |
| | | | 0 |
| CAMA0582.mov | link | SO | -1 |
| | | | -1.5 |
| CAMA0580.mov | link | SO | -2 |
| CAMA0578.mov | link | SO | -3 |
| CAMA0620p.mov | link | PTZ | -3.5 |
| CAMA0576.mov | link | SO | -4 |

# 4.3 Real fire experiments - fireGIS datasets

Before going more into detail on the video smoke analyzer and the spatio-temporal fire risk analysis, this section provides additional information about our real fire experiments. Other examples of certain experiments exist, such as the structural behavior tests of Tisova [14], the Dalmarnock tests [15] in high-rise compartment fires and the multi-compartment Rabot test [16]. However, such experiments are limited in number due to the high costs involved. Numerical simulations, on the other hand, are relatively cheap, but they suffer from computational complexity and the level of detail depends on the time and space discretization. Overall we need to use each opportunity of real fire experiments to further optimize the simulations and use them to get more knowledge in fire behavior.

## 4.3.1 Multi-compartment fire test

**Figure 4.3** An overview of the multi-compartment set-up in association with WFRGent and VIPA (left), smoke visibility risk results mapped on the floor plan of the set-up (right).



A first real large-scale test for evaluation of the platform has been performed in association with WFRGent and VIPA in September 2015. The outcome of these tests was used to investigate the impact of different fire suppression techniques in a multi-compartment set-up, specifically for elderly homes. Previously, elderly homes often had a topology where different, separate rooms are connected to a hallway that is used for circulation and evacuation purposes. Furthermore, common spaces (i.e., living, kitchen) were separated from each other and the spaces were only accessible trough the common hallway. Recently, the design of elderly homes is undergoing a fast evolution. Common spaces become more and more part of the circulation and the evacuation routes and open spaces are becoming mainstream. Despite the increased usage, there are no specific prescriptive rules available for this type of environment and a profound study was needed. The study consisted of a 'sofa fire' experiment, which was performed five times with different conditions:

**First experiment:** Test without suppression system

**Second experiment:** Test with fire and smoke resistant doors

**Third experiment:** Test with a sprinkler installation

**Fourth experiment:** Test with a smoke extraction system

**Fifth experiment:** Test with a combination of sprinkler and smoke extraction

A smoke extraction system, typically referred to as SHC (Smoke and Heat Control system), has the goal to extract the heat and the smoke from a fire. The first reason to install this system is to ensure safe evacuation of the people inside the burning building. Secondly SHC is used to avoid smoke spreading. Finally the mechanism will assist the firefighters during their intervention by providing sufficient visibility to locate the fire source.

A sprinkler installation is a combination of water pipes and sprinkler bulbs. These glass bulbs contain a liquid and break when reaching a specific critical temperature, and the water will be released. The sprinkler system will possibly extinguish the fire, but will mainly limit the fire growth and reduce the temperature of the hot smoke layer.

A watermist system, which is a system with water under pressure that creates very small droplets. The size of the microdroplets depends of the water pressure on the system. This system works at low, medium and high pressure. The higher the pressure, the smaller the water drips and the greater the cooling surface of the drops. For comparison, a normal water drop from a sprinkler system has a diameter of 1 mm. This system seems highly valuable, but it was not taken into account in the fire experiments. The reason to not include the watermist system is that it needs a high pressure connection whereas a standard sprinkler system only needs a low pressure connection. Secondly, watermist systems are a not common fire protection system in Belgium.

For each fire experiment, our fireGIS analyzer generates an overview of the smoke propagation and the smoke thickness. Based on the total amount of smoke in the different scenarios, visualized by the analyzer, it is possible to make objective decisions for new building recommendations. Besides the visual monitoring there were also thermocouple tree measurements of the temperature on several heights and places. With the analysis of Tilley et al. [17] it is possible to determine the height of the neutral plane based on the second derivative of the temperature profile. In general the platform facilitates, in this case, the spatio-temporal understanding of different measures on smoke height, smoke visibility and temperature.

The thermocouple measurements were not performed by our group and therefore they will not be heavily discussed in the thesis. However, some details are given: the temperatures were measured by thermocouples type K, they were able to detect temperatures between 0 and 1000 C. The measurements (14 for each tree) were performed on the following heights, in the fire room: 0,2 m - 0,4m - 0,6 m - 0,8 m - 1,0 m - 1,2 m - 1,4 m - 1,6 m - 1,8 m - 2,0 m - 2,1 m - 2,2 m - 2,3 m en 2,4 m.

Besides the 4 thermocouple trees in the fire room there were also thermocouples placed in:

- The non-fire rooms on 1,9m and 2,4m (to measure the smoke leakage and heating through a fire resistant door)
- At the end of the evacuation routes: above the door and at 2,4m
- Above the fire load (the red couch) at 1,9m and 2,4m

**General conclusions related to the VIPA elderly experiments**

The VIPA studies are organised in collaboration with Gent University, VIPA, Wfrghent. As we were not the head organizers of these tests we were not involved in the design phase of the real fire experiments, however some general conclusions are given below.

First, a compartment is filled very rapidly with smoke. Secondly, there is a pressure built- up in the closed compartments, resulting in the spread of smoke to adjacent compartments separated by fire-resistant doors. Furthermore, the pressure built-up affects the smoke extraction systems negatively. This can be solved by increasing the opening of the room, or by using an automated fire extinguishing mechanism. Thirdly, the smoke passage in a room separated from the combustion chamber by means of two consecutive fire-resistant doors was significantly smaller compared to the one with one door. Finally, a combination of both systems (i.e., smoke extraction and an automated fire extinguishing) was proven to be the best solution for restricting the spread of smoke in the compartment. Therefore, the position of the automatic extinguishing was chosen in such a way that it was not influenced by the airflow obtained by the disruption system.

### 4.3.2 Tunnel fire experiments

The second large-scale fire tests were performed in the Craeybecktunnel. The Craeybeckxtunnel is a tunnel between Brussels and Antwerp (N 51.1005, E 4.2406) in Belgium. In order to investigate the impact of the ventilation system on the

**Figure 4.4** FireGIS experiments at Craeybeckxtunnel (tunnel between Antwerp and Brussels, Belgium).

propagation of the smoke, real pool fire tests were performed by the end of 2014 (Figure 4.4). Besides the monitoring of the visibility metrics and the fire spreading, the recorded video images can also be used for validation of CFD simulations (Figure 4.5), which were performed prior to the tests. It is also important to remark that the ventilation system in the tunnel is transversal to the drive direction. This is not common and gives the opportunity to analyze smoke movement in such circumstances.

**Figure 4.5** Subjective comparison of CFD temperature field (left) and Craeybeckxtunnel video measurements (right). (source CFD calculations FESG)

Prior to the tests, decisions were made related to the fire power. On the one hand, the fire power needed to be limited to avoid severe damage to the tunnel. On the other hand, the power of the fire (see Appendix 1) needs to be realistic to get validable smoke dynamics. In our tests, 20 minute pool fires of 3 MW were generated, which is slightly less then a modern car fire (which is between 4 and 6 MW [18]). In the final test, we also performed a real car fire, with similar results as the pool fires. Different measurements were performed related to temperature, air flow and smoke/visibility. In this Chapter, however, we only focus on the latter sensor type, since only the video sensors were able to monitor the entire tunnel for space-time fire risk analysis. Further explanation of the visibility measurement in fire circumstances is given in the next section.

## 4.4 Video smoke analysis

Video based fire analysis with cameras has been discussed several times in literature over the past years [19]. Different features, such as color, pixel disorder, wavelet analysis are combined with simple linear classifiers or with more advanced deep learning mechanisms. However, the focus in literature was mainly on the detection of the smoke. The propagation of the smoke, the height of the smoke layer or the visibility for example, were investigated less. Furthermore, there is a strong correlation between the visibility, the evacuation movement speed and the surviving chances [20]. To further investigate the visibility we evaluated several video-based visibility metrics in our real fire experiments and developed a quantitative measure that can be used in fire incident management to adapt the tactics of the fire brigades.

The most common features to measure the visibility in images are based on analysis and classification of the brightness, saturation, and contrast pixel values [21]. Additionally, the visibility can also be measured by analyzing the number/strength of visible edges in the image. If these edges are geo-referenced, i.e., labeled with the real location information, it is also possible to say how far it is possible to see. Narvekar et al. [22], for example, use this methodology to measure the sharpness of an image. If the number of edges in a particular image block is higher than a pre-defined threshold value, the block is labeled ”high visibility”. In our work, the opposite approach could be used to detect a decrease in visibility, i.e., smoke increase. In order to use each of these techniques some video-based training of the environment is needed, as explained by Hassanpour et al. [23].

### 4.4.1 Low-cost video smoke analyzer

A flowchart of our low-cost algorithm for video smoke analysis is shown in Figure 4.6. The algorithm starts by converting the video to HSV color space [24] and by filtering out the value (V) component. In this way, a change in lightning or a change in colors will not affect the algorithm [19]. Next, the Canny edge detector [25] is used to detect the prominent edges in V. This edge detector uses Gaussian filtering and hysteresis tracking to smoothen the image, remove the noise, and to suppress the weakly connected edges (See Appendix B.1.5 for more technical details). Subsequently, we count the remaining bright pixels in the upper part of the image. This value gives a quantitative measure for the visibility in that region, i.e., an indication of the smoke level. We only focus on the upper part of the images, since moving objects (like people and cars) in the lower part of the image can disturb the algorithm. Furthermore, smoke will rise, thus the upper part will contain most of the smoke. Finally, we normalize the edge counts (using edge characteristics of the video training phase) and we calculate the smoke risk level $L$

ranging from 1 to 5, i.e., high visibility and no-visibility respectively (see Figure 4.7). As the smoke risk level is a ratio there is no unit.

**Figure 4.6** Flowchart of the low-cost video smoke analyzing algorithm.

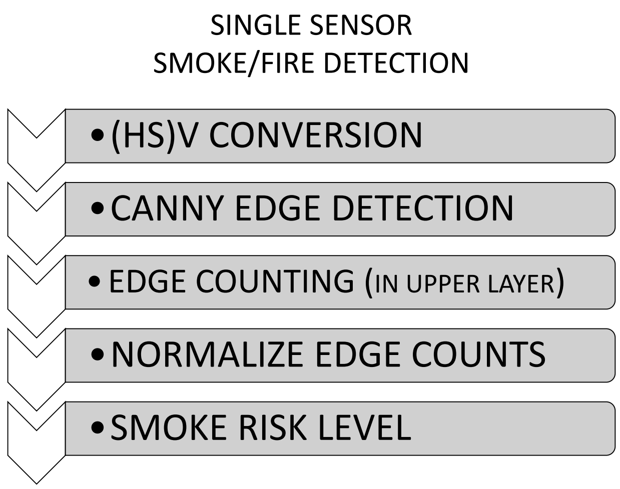



**Figure 4.7** Indicative smoke risk levels (left), their corresponding visibility level (middle) and the impact on the evacuation according to Jeon et al. [20] (right).

| Smoke risk level | Visibility | Influence on the evacuation |
|---|---|---|
| 1 | High visibility | Unaffected |
| 2 | Moderate visibility | Slightly affected |
| 3 | Reduced visibility | Moderate affected |
| 4 | Strongly reduced visibility | Severely affected |
| 5 | No-visibility | Disorientation |

It is important to remark that all these operations have a low computational cost (i.e., the time complexity to calculate the edges on a m x n image is equal to O(mn log mn) according to the Big-O complexity analysis [26]), making it possible to process the video frames in real-time on a standard processor. After the visibility estimation, the resulting smoke risk levels are stored in a comma-separated values (CSV) file, as shown in Figure 4.8. For each sensor that is used in the fire experiments, we generate a comma-separated object containing the position (latitude/longitude coordinates from the sensor meta data) and the smoke risk level *L* at time-stamp *T*. Based on this CSV file, fire maps can be generated. Besides the CSV output it is also possible to transform the data to the standard geojson format [27], making it easy to reuse the data in a GIS application. The geojson format is a specification describing the encoding of a variety of geographic data structures. GeoJSON supports the following geometry types: Point, Lines, Polygon and their combinations.

**Figure 4.8** CSV files with detected smoke risk levels. For each timestamp *T*, the coordinates of the cameras and corresponding risk levels *L* are stored in comma-separated objects.

dataCAMS2.js ×

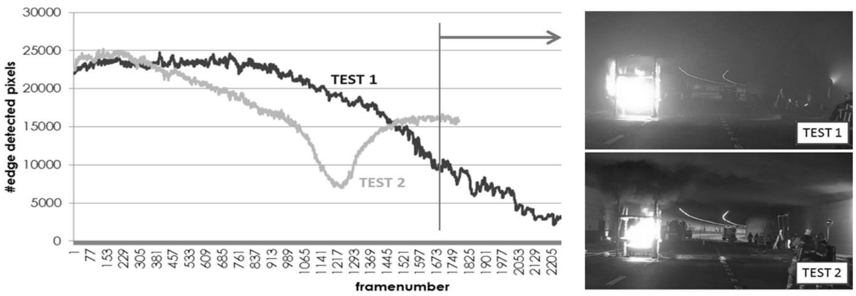
```
var T164= [[51.185478,4.415953,1.1968],[51.184678,4.416353,3.6634]
var T165= [[51.185478,4.415953,1.1078],[51.184678,4.416353,3.678],
var T166= [[51.185478,4.415953,1.1947],[51.184678,4.416353,3.6489]
var T167= [[51.185478,4.415953,1.1422],[51.184678,4.416353,3.623],
var T168= [[51.185478,4.415953,1.1283],[51.184678,4.416353,3.6383]
var T169= [[51.185478,4.415953,1.1035],[51.184678,4.416353,3.6282]
var T170= [[51.185478,4.415953,1.1284],[51.184678,4.416353,3.6359]
var T171= [[51.185478,4.415953,1.1029],[51.184678,4.416353,3.587],
var T172= [[51.185478,4.415953,1.1691],[51.184678,4.416353,3.5362]
```

### 4.4.2 Evaluation of the smoke analyzer

First of all, an evaluation of the estimated visibility levels of the proposed computational low-cost video smoke analyzing algorithm is done by qualitative analysis of the video fragments of the real fire experiments (see Figure 4.9). Temporal analysis on the results of the output of the algorithm for our real tunnel fire experiments shows that changes in visibility are properly detected. Figure 4.10 indicates that a decrease and a clearing of Test 2 is correctly estimated. In Test 1 we encountered a continuous decrease in visibility, which also corresponds to the output of our algorithm.

**Figure 4.9** Combined video images for subjective evaluation of Craeybeckxtunnel experiments.

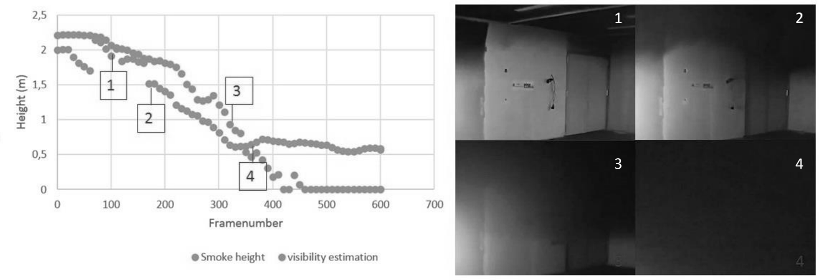

**Figure 4.10** Comparison between the output of the proposed smoke analyzer algorithm (shown left) and the visual results of real fire experiments (shown right).




A second evaluation, shown in Figure 4.11, compares the estimated smoke layer height from the thermocouple trees (shown in blue) and the visibility estimations generated by the algorithm (shown in orange). If the height of the smoke interface layer is lower than the vertical position of the camera there is a certainty that there must be a strong reduction in the visibility.

We estimated the height of the smoke layer with thermocouple trees (i.e., a stacked array of thermocouples to measure the temperature at different heights) in the enclosure set-up with two methodologies of He et al. [28]. The first method, the critical temperature method, investigates the temperature profile and the point where the temperature rise, relative to its initial temperature exceeds a given threshold is taken as the smoke layer. The second method investigates the relative rise in the temperature profile. The height where the temperature rise exceeds N percent of the maximum temperature is taken as the smoke layer height. For label 4 in Figure 4.11, for example, the averaged thermocouple based estimation of the smoke height is 0.6 m. This is lower than the position of the camera, which was placed on 0.8 m. The result of the camera based visibility estimation also corresponds to the subjective evaluation of the image at the time corresponding with label 4 (shown right). From the smoke layer estimation height and the reduced visibility we can confirm the statement: *If the height of the smoke interface layer is lower than the vertical position of the camera there is a certainty that there must be a strong reduction in the visibility.*

**Figure 4.11** Comparison between the estimated smoke layer height from the thermocouple trees (shown in blue) and the camera based visibility estimations (shown in orange). Corresponding images for certain moments in time (label 1-4) are shown in the picture right.




### Cost and performance analysis

A performance comparison between thermocouple trees and visual sensors reveals that a camera solution will fail already in case the sensor is located beneath the smoke layer. However, thermocouples are point measurements and not omnipresent in exisiting buildings.

On the algorithm side, the computational cost of the smoke layer estimation algorithm is O(n) as it just needs to find the minimum value. The total cost of the visibility algorithm given an input image with dimensions mxn is O (mn log mn).

## 4.5 Intra and inter variance of the temporal analysis

In order to generate visibility levels more accurately, i.e., smoke risk estimations, we propose to combine the detection results of each of the single-view sensors and analyze them together in a multi-sensor analysis set-up. Previous research has illustrated that a multi-sensor data fusion approach improved the single sensor decision process for fault detection [29, 30]. Since we have multiple sensors monitoring the scene from different viewpoints, problems that arise in one sensor can be compensated by the other surrounding sensors. Further explanation is done for the camera and smoke analyzing system, but the approach can also be used for other types of sensors.

Based on the estimated smoke risk level *L* of all the camera views, we propose a multi-view extension of the single-view algorithm based on the within- and between-variance of the visibility estimations. By analyzing the differences between the *L* values, an indication on the accuracy of the measurements at each moment *T* in time can be given. For example, if the variance $\sigma_w(L, v_i)$ (Eq. 4.1) of $L_T$ values over time is high in one of the camera views $v_i$, it is less safe to trust the measurements of this view compared to a camera view $v_j$ in which the within-variance of *L* is smaller. Furthermore, it is also important to check the between-variance $\sigma_b(L)$ (Eq. 4.2) of neighboring camera views, since this can give an indication regarding the certainty of the measurements. When large differences are observed, i.e., $\sigma_b(L)$ is high, this can be an indication that it is not safe anymore to trust the cameras and it is better to rely on other types of sensors. In contrast, when neighboring sensor estimations closely follow each other, this indicates that their estimations can be trusted with high(er) probability. Similar metrics have been proposed and evaluated in the multi-compartment full-scale 'Rabot' fire tests [9] [8].

$$\sigma_w(L, v_i) = \frac{1}{N}\sum_{T=1}^{N}(L_T(v_i) - \overline{L_{1..N}(v_i)})^2 \tag{4.1}$$

$$\sigma_b(L) = \frac{1}{N}\sum_{j=1}^{M}(\overline{L(v_j)} - \overline{\overline{L(v_{1..M})}})^2 \tag{4.2}$$

Using the above mentioned within- and between-variance, the estimated accuracy of each sensor type can be monitored and the most reliable sensors can be selected at each moment in time. The impact of the proposed system can be seen in Figure 4.12, which shows an example with 3 'correct' sensors and one sensor giving wrong visibility levels (due to the interference of the video signal with the vibrations generated by the smoke extraction system). It is important to remark that the calculation of the nearby sensor will only work in case the video sensors share a common field of view. If the sensor is placed behind a smoke resistant door, the standard variation will be high, but this is not due to a failed camera. The proposed algorithm (error calculation for neighbor sensors) indicates this error correctly, i.e., the standard variation with the wrong sensor was 0.01389 whereas the value without the broken sensor was very low (0.000261). The impact of the broken sensor can also be seen in the heatmap (see Figure 4.13). With the faulty sensor included, the red zone (no-visibility) is smaller compared to the correct representation.

**Figure 4.12** Smoke visibility levels of 3 correct neighboring sensors and one broken/faulty sensor (due to the interference with the smoke extraction system).

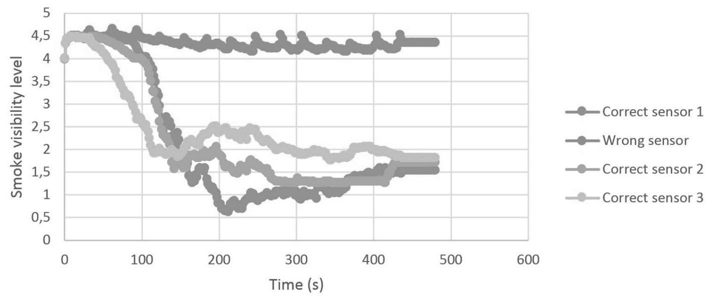
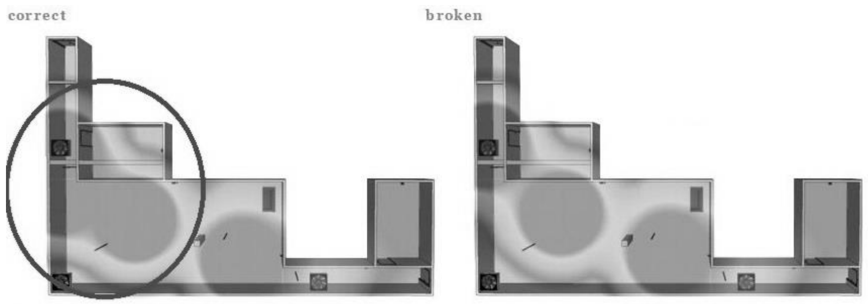
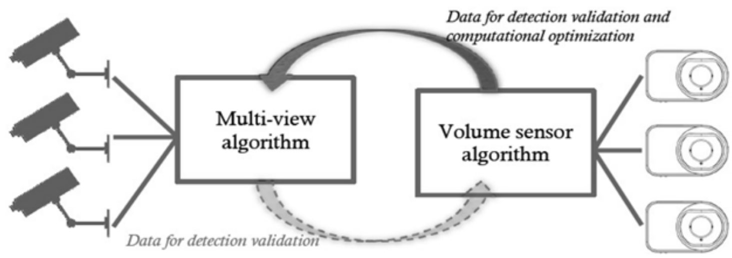



**Figure 4.13** Heatmap representation with the corrected values (left) and with one broken sensor (right).




Some other methods exist to analyze the smoke variance [30], for example, Gonzales et al. investigated the gain achieved by combining volume sensors, like the Xtralis OSID beam sensors (an Open-area Smoke Imaging Detection beam sensor consists of a transmitter and receiver of infrared and UV light. The strength of both signals is compared to calculate the smoke presence or absence.) [31], and Video Fire Detection (VFD) technologies. By using sensors that do not depend on illumination and/or are immune to steam and dust, the VFD detection problems can be addressed directly. This multi-sensor/multi-modal approach leads to improved smoke alarm verification, smoke localization validation and smoke layer depth measurement. A representation of the multi-sensor combination and the algorithms interactions can be seen in Figure. 4.14. By integrating the output from OSID with the information from VFD, both detection systems share information (i.e., give feedback to each other). Due to the generic architecture of fireGIS, similar sensor combinations can easily be integrated in the proposed architecture. An example of this incorporation has already been given by the assimilation of thermocouple trees. The selection of the specific redundancy is depending on the use case. In case of the tunnel experiments, the inter variance could easily indicate if one sensor failed, but in case of the VIPA study more redundancy should be built in as one camera per room is not reliable enough. Subsequently, there should be redundancy in the sensors (i.e., different types of sensors measuring the same values), but also redundancy in the connection (e.g., Wifi, wired, Bluetooth LE) and finally an additional layer with anomaly detection that can indicate wrong sensor values or problems in the system.

**Figure 4.14** Algorithms interaction of video and OSID fire detection sensors [30].




## 4.6 Video flame analysis

Besides the analysis of the smoke visibility, the smoke spreading and the smoke interface layer, the flame height is a valuable source of information to understand the fire behavior. The mean flame height is theoretically described by Heskestad [32] as the level where the combustion reactions are essentially complete. Furthermore, flame dimensions can give valuable information about the Heat Release Rate (HRR) of a fire, as proposed by Beji et al. [9]. To eventually measure the height there are different techniques and algorithms proposed in literature.

- A standard method to get the flame height is to analyze the temperature profile measured central in the fire (by robust thermocouples). The characteristic mean temperature rise of 500 - 600 degrees is taken in the center of the flame.

- Ukleja et al. [33] described the flame height as the position where the flame presence probability is higher than 50 percent over time.
- Verstockt et al. [34] proposed a spatio-temporal analysis of the flame height, by first applying automatic thresholding on the bright (flame) pixels and subsequently applying a low-cost hand crafted flame feature analysis (bounding box disorder and Otsu thresholding).
- More recently CNN architectures (as introduced in Chapter 3 and explained more thoroughly in Appendix 2) are exploited to detect the flame region from a video input [35]. Furthermore, segmentation networks, such as Squeezenet, achieve fine grained results on input images for fire region segmentation and eventually fire shape analysis [36].

### 4.6.1 Flame height algorithm

To perform a spatio-temporal flame dimension analysis of real fire experiments we propose a novel algorithm that combines temporal slicing and SLIC segmentation. The usability of the proposed algorithm has been evaluated during the experimental study of corner fires, performed by Zeinali et al. [37] and Beji et al. [38]

#### Temporal slicing

Temporal slicing is a common technique in video summarization [39] and motion estimation [40]. Figure 4.15 gives a graphical representation and visualization of the slicing methodology on a couch fire video footage performed by Exova[1]. The video sequence is a 3D volume with x and y the image dimensions and time as temporal dimension. The slice is a temporal collection of strips (x,t) or (y,t), a strip is a row (vertical slice) or column (horizontal slice) from the image frame that are combined to one image. On that one particular image we can easily interpret and examine the temporal evolution (i.e., the change in width or height of the flame front) of that row or column.

#### Superpixels SLIC

To segment the fire region and to estimate the fire dimensions (from the temporal slice images) we used the Simple Linear Iterative Clustering (SLIC) segmentation algorithm [41]. Initially K (i.e., on average 200 clusters over the complete image) regularly spaced cluster centers are selected. Secondly, each pixel is associated to the nearest cluster center according to the proximity and the color similarity in the CIELAB color space. Thirdly, the pixels are iteratively associated to a cluster and the cluster centers are recomputed up to convergence.

[1] `https://www.exova.com/`

**Figure 4.15** Temporal slicing for couch fire experiment performed in Exova, left the schematic overview of the slices, upper right the horizontal slice, down right the vertical slice.

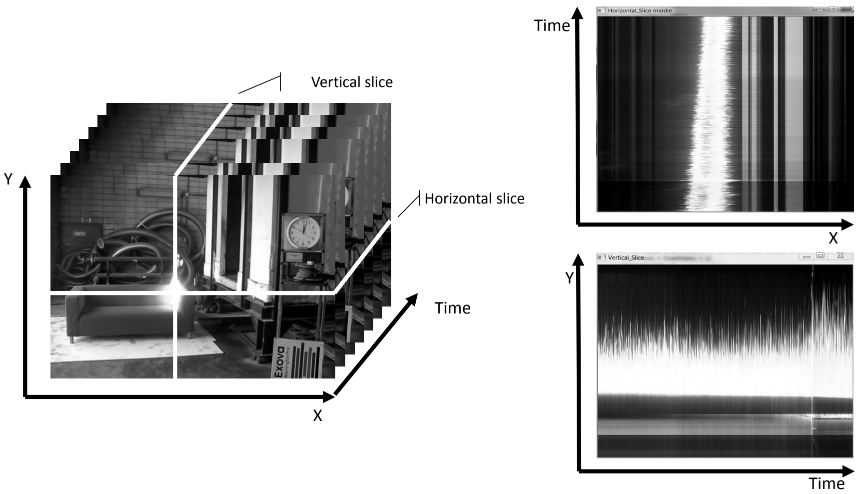
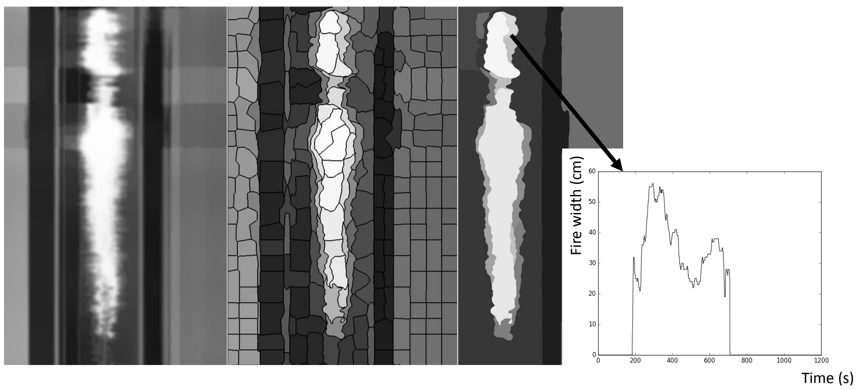



**Figure 4.16** Schematic overview of the video flame analysis, temporal slice (left), SLIC segmented region (middle) and merged clusters and corresponding height profile (right).




## Algorithm

Figure 4.16 gives a schematic overview of the proposed flame analysis tool. Firstly the temporal slicing mechanism is used to extract a temporal overview of the flame interaction at a certain position. Currently the slice position is selected manually and the algorithm is optimized for fixed cameras. Changing the viewpoint, the lightning conditions or occluding the camera would affect the temporal slice result. However, with global motion detection it should be possible to determine the new position of the slice for non-fixed cameras. Furthermore, in case of a viewpoint-change a fire detection algorithm could be used to detect the largest fire front. Secondly, the SLIC superpixel mechanisms segments fire pixels in similar regions (color and intensity). Thirdly the clusters are transformed to a region adjacency graph [42] where each node in the graph represents a cluster with a set of pixels. The weight between two adjacent regions represents how similar or dissimilar the clusters are. Subsequently, the clusters are merged until only highly similar region pairs remain. Furthermore, the fire height (i.e., the horizontal pixel coordinates of the brightest or selected cluster) is determined and a mapping between the pixel distance and the real world coordinate distance is implemented. Finally, by plotting for each time slice (row in the image) the height of the fire, the final height profile is created.

### 4.6.2 Video puffing and flickering frequency

Besides the flame height, the flickering frequency is a quantitative measure to describe the flame. A typical flame will have a non-constant flame height due to vortex shedding and air entrainment. For typical enclosure elements (e.g., furniture fires) the pulsation frequencies are between 0.4 and 10 Hz [43]. Furthermore, the frequency and the height depend on the 'fuel-air' ratio, the temperature, the flame diameter and the room configuration, as stated by Bhaduri et al. [43]. A frequently used calculation method is the algorithm of Zukoski et al. [44] where the puffing or flickering frequency (f in Hz) of a flame (free burning pool plume) is described by the normalized pool diameter (D in m) and the gravitational acceleration (g = 9.81 m/s$^2$):

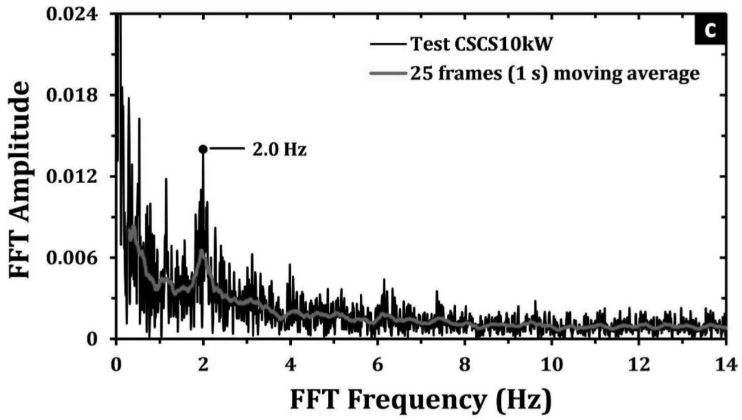
$$f = (0.5 \pm 0.04).(\frac{g}{D})^{0.5} \tag{4.3}$$

In our video flame analysis tool the puffing frequency is empirically estimated by applying a Fast Fourier analysis on the flame height estimations given by the flame height algorithm as proposed in Section 4.6.1. According to the Nyquist theorem:("*A sufficient sample rate is at least two times the highest expected frequency.*") the sampling rate is adapted to the expected frequency (e.g., between 0.4 and 10 Hz). Figure 4.17 indicates the puffing frequency of 2Hz in the corner fire experiments of Zeinnali et al. [37].

## 4.7 Smoke map generation

We developed an innovative distributed system that combines regular cameras with geographic information web mapping technology. In order to generate a 2D smoke map of the smoke risk levels at time-stamp *T*, we developed a dynamic JavaScript-based web page. The web page makes use of the Leaflet.heat and leaflet.js heatmap

**Figure 4.17** Fast Fourier analysis on the flame height curve. Notify the clear peak (significant frequency) around 2Hz (puffing frequency).




plugin, which is a simple and fast solution for heatmap generation. This plugin[2] constructs a heatmap layer on top of a map or an image, given an array of latitude/longitude points and intensity, i.e., the corresponding smoke risk level. Figure 4.18 shows an example of two temporal smoke maps generated for one of the Craeybeckxtunnel tests using the heatmap functionality. At t=80s after ignition (shown left), only the small central part of the tunnel has low visibility (red color), while the other parts of the tunnel are still smoke-free (blue zones). 10 seconds later, smoke starts spreading towards both sides of the tunnel (as shown in the map on the right), indicating low visibility over the entire tunnel. This kind of spatial and temporal information can be very useful for fire incident management, such as evacuation planning. Similary, Figure 4.19 shows a selection of smoke maps generated for the second, third and fifth multi-compartment test (see Section 4.3) at different moments in time (T=200s and T=500s). With these heatmaps it is easier to compare different set-ups and their corresponding visibility. This straightforward comparison will assist decision makers in defining new rules for fire prevention. Without these maps, a manual, subjective evaluation of different videostreams should be done in combination with CFD numerical experiments. Such manual verifications would require a lot of manual labour and would be more error-prone. Small changes in the visibility would remain unnoticed and getting an overview of several viewpoints would also be very difficult.

[2] `http://leafletjs.com.`

**Figure 4.18** FireGIS heatmaps showing temporal evolution of smoke risk level (i.e., low visibility) in the Craeybeckxtunnel experiment.

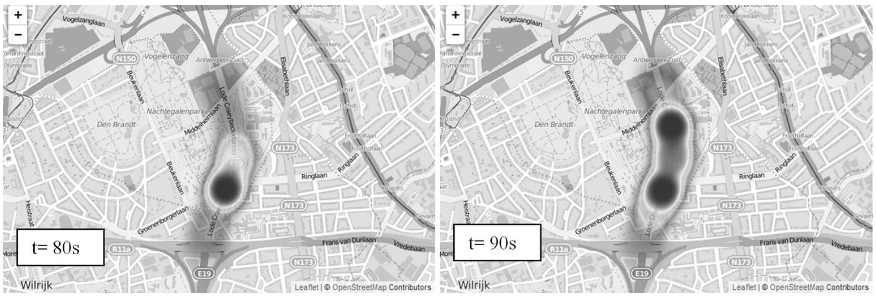
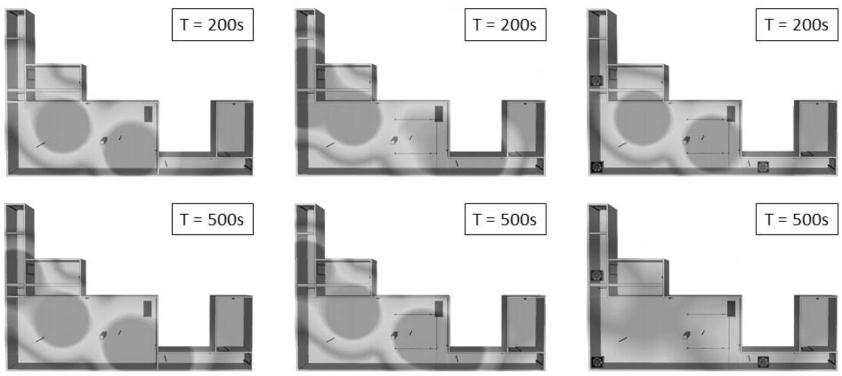



**Figure 4.19** FireGIS heatmaps showing the temporal evolution of the smoke risk level in the multi-compartment experiments. The left, middle and right images show the temporal evolution at T=200s and T=500s for the second, third and fifth test (described in Section 4.3), respectively.

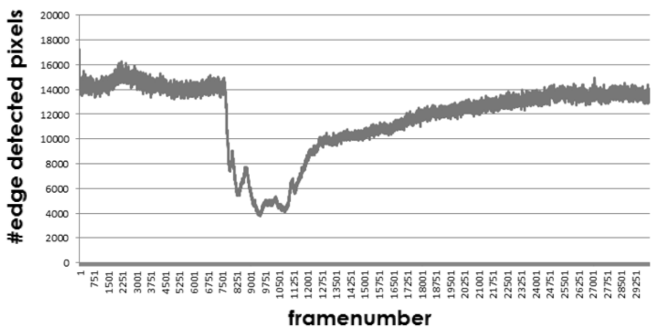



# 4.8 Spatio-temporal fire risk analysis

## 4.8.1 Spatio-temporal visibility analysis

By analyzing the smoke maps (shown in Figure 4.18 and 4.19) over time, it is possible to perform a space-time analysis of the smoke spreading and to get an idea about the direction, speed and thickness of the smoke at each point in time. This can facilitate the smoke reading and decision making, as discussed in the introduction (see Chapter 1) of this thesis. Using the CSV smoke risk data, the fireGIS platform can also plot temporal graphs of the smoke risk level (edge counts) for a specific sensor region. This facilitates the understanding of the temporal changes of the visibility for a certain test and a certain sensor. Figure 4.20 illustrates this process. The plot visualizes a drop followed by a clearing of the visibility for the sensor in the middle of the tunnel experiment. Similar trends/evolutions can also

**Figure 4.20** Temporal evolution of edge counts in the middle of the tunnel. First there is a good visibility (high amount of edge pixels), after 5000 frames there is a strong reduction (i.e., a quarter of the edge pixels) and after 12 000 frames there is a clarification of the visibility (amount of pixels).




be detected by subjectively analyzing the combined, i.e., stitched, video images in Figure 4.21. However, objective results, as those shown on the temporal smoke risk graph (Figure 4.20), are easier and faster to interpret compared to video images. The video streams can obviously help in the evaluation of post-fire analysis tasks. When smoke deposition on the lens, or dirt or other particles on the camera,

**Figure 4.21** Combined video images for subjective evaluation of Craeybeckxtunnel experiments.


the values of the smoke risk data will be distorted. The visibility for instance will be much lower when there is fog on the lens. This can be solved in two ways. The first method is to incorporate another sensor (such as a beam sensor which is less sensitive to fog) and take the normative value between both type of sensors [45]. The second method is to use lenses with self-cleaning films such as proposed by Thompson [46].

### 4.8.2 Spatio-temporal temperature mapping

In addition to the video smoke risk levels, the fireGIS platform also gives the possibility to generate spatio-temporal temperature graphs for each sensor region based on the available thermocouple trees. Figures 4.22 and 4.24 show the temporal evolution of the temperature for two different tests performed in the multi-compartment set-up. Compared to the traditional representation of the values of the corresponding thermocouple trees (shown in Figures 4.23 and 4.25) this type of visualization is easier to get a better understanding of temperature changes over time and to analyze the height changes. Preliminary user experience evaluations with members of the fire brigade and the governmental agency have shown that this new type of visualization could give more insight into different set-ups and that this facilitates faster understanding of real fire tests. In the next chapter, we will discuss the subjective evaluation tests to get improved selection and visualization criteria.

**Figure 4.22** Temporal evolution of the temperature close to the fire, in a set-up without a sprinkler system. The horizontal axis shows the time and the vertical axis shows the height of the thermocouple measurement. The colors represent the temperature (blue is cold, red is hot).

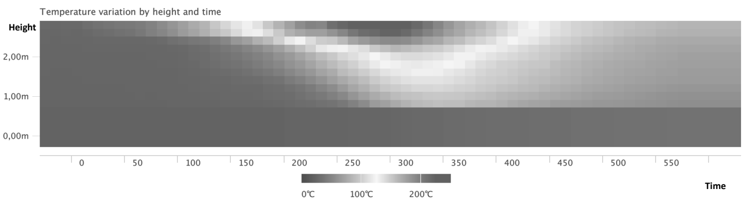



**Figure 4.23** Traditional way of representing temporal evolution of thermocouple values for the set-up without a sprinkler system. The temperature (on the vertical axes) changes over the time (horizontal axis).

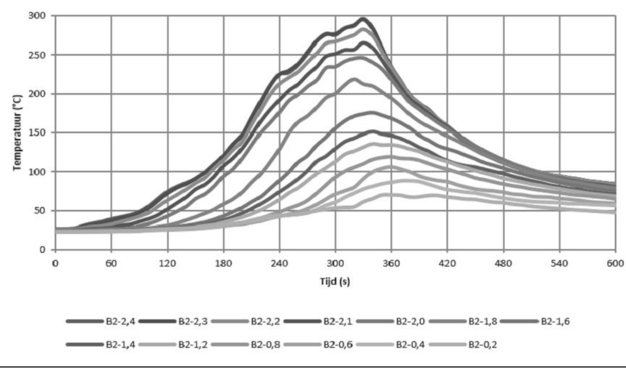


**Figure 4.24** Temporal evolution of the temperature close to the fire, in a set-up with a sprinkler system. The horizontal axis shows the time and the vertical axis shows the height of the thermocouple measurement. The colors represent the temperature (blue is cold, red is hot). The red line indicates the activation time of the sprinkler system (240s after fire ignition).

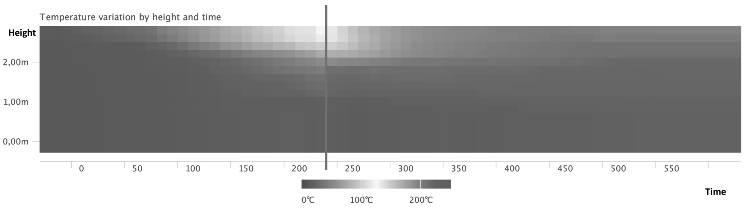
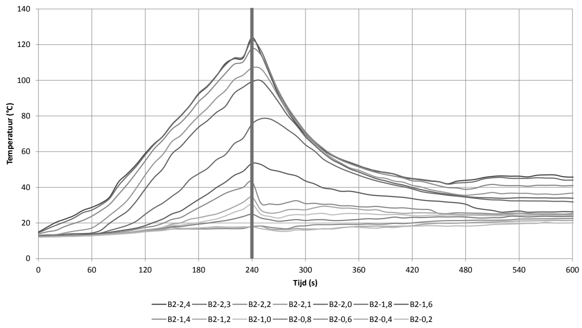



**Figure 4.25** Traditional way of representing temporal evolution of thermocouple values for the set-up with a sprinkler system. Each line represents a thermocouple on a certain height (see color labels). The temperature (on the vertical axes) changes over the time (horizontal axis). The red line indicates the ignition time of the sprinkler system (240s after fire ignition).



## 4.9 Conclusions and future work

This chapter presented the generic architecture of the fireGIS framework, which allows the generation of real-time heatmaps that show the space-time distribution of fire risk levels and fire behaviour characteristics. In order to show the feasibility of the proposed platform, real-fire experiments have been performed in a large-scale road tunnel and in a multi-compartment set-up. Video sensors and thermocouples have been used as input to feed the fireGIS system, and the visibility-based video fire analyzing results are mapped to spatio-temporal heatmaps. Those maps can assist decision makers in taking actions and can facilitate quick fire emergency response. Finally, user evaluations have already shown that the platform is easy to understand and that it improves the current visualization of different fire characteristics (see Figure 4.23).

# References


[1] Chengcheng Gai, Wenguo Weng, and Hongyong Yuan. *GIS-based forest fire risk assessment and mapping*. In Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on, pages 1240–1244. IEEE, 2011.

[2] Kuno Kurzhals and Daniel Weiskopf. *Space-time visual analytics of eye-tracking data for dynamic stimuli*. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2129–2138, 2013.

[3] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. *Space, time and visual analytics*. International Journal of Geographical Information Science, 24(10):1577–1600, 2010.

[4] David W Dodson. *The art of reading smoke*. PennWell Corporation, 2007.

[5] Lars-Göran Bengtsson. *Enclosure fires*. Swedish Rescue Services Agency, 2001.

[6] Baaij S Lambert K. *Brandverloop technisch bekeken en tactisch toegepast*. SDU uitgevers, 2011.

[7] Steven Verstockt. *Multi-modal video analysis for early fire detection*. PhD thesis, Ghent University, 2011.

[8] Steven Verstockt, Sofie Van Hoecke, Tarek Beji, Bart Merci, Benedict Gouverneur, A Enis Cetin, Pieterjan De Potter, and Rik Van de Walle. *A multi-modal video analysis approach for car park fire detection*. Fire Safety Journal, 57:44–57, 2013.

[9] Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. *On the use of real-time video to forecast fire growth in enclosures*. Fire Technology, 50(4):1021–1040, 2014.

[10] Michael Compton, Payam Barnaghi, Luis Bermudez, RaúL GarcíA-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, et al. *The SSN ontology of the W3C semantic sensor network incubator group*. Web semantics: science, services and agents on the World Wide Web, 17:25–32, 2012.

[11] John Adedapo Ojo and Jamiu Alabi Oladosu. *Video-based Smoke Detection Algorithms: A Chronological Survey*. Computer Engineering and Intelligent Systems, 5(7):38–50, 2014.

[12] Wei Ye, Jianhui Zhao, Song Wang, Yong Wang, Dengyi Zhang, and Zhiyong Yuan. *Dynamic texture based smoke detection using Surfacelet transform and HMT model*. Fire Safety Journal, 73:91–101, 2015.

[13] A Enis Cetin, Bart Merci, Osman Günay, Behçet Ugur Töreyin, and Steven Verstockt. *Methods and Techniques for Fire Detection: Signal, Image and Video Processing Perspectives*. Academic Press, 2016.

[14] Jakob Degler, Andreas Eliassion, Johan Anderson, David Lange, and David Rush. *A-priori modelling of the Tisova fire test as input to the experimental work*. In The First International Conference on Structural Safety under Fire and Blast, Volume: 1, 2015.

[15] Cecilia Abecassis-Empis, Pedro Reszka, Thomas Steinhaus, Adam Cowlard, Hubert Biteau, Stephen Welch, Guillermo Rein, and Jose L Torero. *Characterisation of Dalmarnock fire test one*. Experimental Thermal and Fluid Science, 32(7):1334–1343, 2008.

[16] Steven Verstockt, Tarek Beji, Bart Merci, and Rik Van de Walle. *RABOT2012: presentation of a multi-view video dataset of the full-scale ('Rabot') fire test*. In 7th International Seminar on Fire and Explosion Hazards, pages 58–66. Research Publishing, 2013.

[17] Nele Tilley, Pieter Rauwoens, and Bart Merci. *Verification of the accuracy of CFD simulations in small-scale tunnel and atrium fire configurations*. Fire Safety Journal, 46(4):186–193, 2011.

[18] Bart Merci and M Shipp. *Smoke and heat control for fires in large car parks: Lessons learnt from research?* Fire safety journal, 57:3–10, 2013.

[19] A Enis Çetin, Kosmas Dimitropoulos, Benedict Gouverneur, Nikos Grammalidis, Osman Günay, Y Hakan Habibolu, B Uur Töreyin, and Steven Verstockt. *Video fire detection–Review*. Digital Signal Processing, 23(6):1827–1843, 2013.

[20] Gyu-Yeob Jeon, Ju-Young Kim, Won-Hwa Hong, and Godfried Augenbroe. *Evacuation performance of individuals in different visibility conditions*. Building and Environment, 46(5):1094–1103, 2011.

[21] Martin Roser and Frank Moosmann. *Classification of weather situations on single color images*. In Intelligent Vehicles Symposium, 2008 IEEE, pages 798–803. IEEE, 2008.

[22] Niranjan D Narvekar and Lina J Karam. *A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection*. In

Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on, pages 87–91. IEEE, 2009.

[23] Hamid Hassanpour, Mehdi Sedighi, Ali Reza Manashty, et al. *Video frames background modeling: Reviewing the techniques*. Journal of Signal and Information Processing, 2(02):72, 2011.

[24] Rafael C Gonzalez, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.

[25] John Canny. *A computational approach to edge detection*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, (6):679–698, 1986.

[26] Samy S Abu Naser. *Big O Notation for Measuring Expert Systems complexity*. 1999.

[27] Howard Butler, Martin Daly, Allan Doyle, Sean Gillies, Tim Schaub, and Christopher Schmidt. *The GeoJSON format specification*. Rapport technique, 67, 2008.

[28] Yaping He, Anthony Fernando, and Mingchun Luo. *Determination of interface height from measured parameter profile in enclosure fire experiment*. Fire Safety Journal, 31(1):19–38, 1998.

[29] Rolf Isermann. *Model-based fault-detection and diagnosis–status and applications*. Annual Reviews in control, 29(1):71–85, 2005.

[30] Steven Verstockt, Rik Van de Walle, Luis Alejandro Gonzalez Avila, Bart Merci, and John De Blonde. *Combining volume sensors with multi-modal video analysis for fire detection and forecasting*. In International Conference on Automatic Fire Detection, volume 2, pages 203–210, 2014.

[31] Ron Knox. *Open-Area Smoke Imaging Detection (OSID).*

[32] Morgan J Hurley, Daniel T Gottuk, John R Hall Jr, Kazunori Harada, Erica D Kuligowski, Milosh Puchovsky, John M Watts Jr, CHRISTOPHER J WIECZOREK, et al. *SFPE handbook of fire protection engineering*. Springer, 2015.

[33] MATHIEU Suzanne, SEBASTIAN Ukleja, MA Delichatsios, JIANPING Zhang, and BJORN Karlsson. *Fundamental flame spread and toxicity evaluation of fire retarded polymers*. Fire Safety Science, 11:846–859, 2014.

[34] Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J Hollemeersch, and Rik Van de Walle. *FireCube: a multi-view localization framework for 3D fire analysis*. Fire Safety Journal, 46(5):262–275, 2011.

[35] Zhen Zhong, Minjuan Wang, Yukun Shi, and Wanlin Gao. *A convolutional neural network-based flame detection method in video sequence*. Signal, Image and Video Processing, pages 1–9, 2018.

[36] Khan Muhammad, Jamil Ahmad, Zhihan Lv, Paolo Bellavista, Po Yang, and Sung Wook Baik. *Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, (99):1–16, 2018.

[37] Davood Zeinali, Steven Verstockt, Tarek Beji, Georgios Maragkos, Joris Degroote, and Bart Merci. *Experimental study of corner fires : part I : Inert panel tests*. COMBUSTION AND FLAME, 189:472–490, 2018.

[38] Tarek Beji, Bart Merci, Steven Verstockt, and Pascal Zavaleta. *Flame spread monitoring and estimation of the heat release rate from a cable tray fire using Video Fire Analysis (VFA)*. FIRE TECHNOLOGY, 52(3):611–621, 2016.

[39] Chong-Wah Ngo, Ting-Chuen Pong, and Roland T Chin. *Video partitioning by temporal slice coherency*. IEEE Transactions on Circuits and Systems for Video Technology, 11(8):941–953, 2001.

[40] S-L Peng. *Temporal slice analysis of image sequences*. In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, pages 283–288. IEEE, 1991.

[41] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. *Slic superpixels*. Technical report, 2010.

[42] Alain Trémeau and Philippe Colantoni. *Regions adjacency graph applied to color image segmentation*. IEEE Transactions on image processing, 9(4):735–744, 2000.

[43] AA Putnam, FE Belles, and JAC Kentfield. *Pulse combustion*. Progress in energy and combustion science, 12(1):43–79, 1986.

[44] Edward E Zukoski, Toshi Kubota, and Baki Cetegen. *Entrainment in fire plumes*. Fire safety journal, 3(3):107–121, 1981.

[45] A Enis Çetin, Bart Merci, Osman Günay, Behçet Ugur Töreyin, and Steven Verstockt. *Methods and techniques for fire detection: signal, image and video processing perspectives*. Academic Press, 2016.

[46] CS Thompson, RA Fleming, and M Zou. *Transparent self-cleaning and antifogging silica nanoparticle films*. Solar Energy Materials and Solar Cells, 115:108–113, 2013.

# 5

# Fireground understanding and data visualization

*This chapter reports on the added value of combining different types of sensor data and geographic information for fire incident management. A survey was launched within the Belgian fire community to explore the necessity and the use of new types of sensor data during a fire incident. Furthermore, an evaluation of different visualization set-ups was done using mock-up screens to gain further insights in fire data reporting. This evaluation revealed that people are visually-oriented and that video footages and images are of great value to gain insights in a particular problem. However, due to the limited available time (i.e., fast decisions need to be taken) and the large amount of cameras it is not feasible to analyze all video footages sequentially. To solve this problem we propose a video summarization mechanism and a video highlight selection tool.*

## 5.1 Introduction

During the first minutes of an intervention, the incident officer needs to take fast and appropriate decisions with very limited information. However, the scene layout, the structure of the building, the fire dimensions and position, the hazardous materials present, and the status and the position of victims, are factors that affect the initial strategy and tactics. In previous chapters, mechanisms have been proposed for scene understanding, fire characteristic estimation and building analysis, but to have a fast overview of the situation, incident management tools are necessary.

Currently, there are two main communication tools for fire incident management: a broadcast channel for voice communication and a second textual channel for status updating. However, from a psychological perspective it is evaluated that our brain is better at processing visuals than text [1]. Our brain can process an image 60 000 times faster than a piece of text. On average, when a person sees a particular image for the first time, it takes 113 milliseconds to process the information. In that sense it is advisable to stress the development of image related tools to facilitate the incident officers instead of improving the existing auditive or textual communication tools that have problems with reliability (e.g., connection-loss in concrete buildings) and interoperability.

To generate the needed visual information we can exploit the increased performance and quality of fixed security cameras and handheld devices such as smartphones and tablets. It is important, however, that the video footage is captured in the most appropriate way for visualization to a fire incident manager. For this reason, guidelines should be given for handheld camera handling during an intervention (see Section 6.3). Subsequently, the sensor input of the fixed and handheld cameras can be streamed to the incident commander to have an overview of the incident scene. However, in Section 5.2 we will show that it is not possible for a person to analyze all the video streams in a sequential manner due to the cognitive disability. There is a need for a video summarization mechanism that filters out the redundant and unnecessary frames, while preserving the distinctive frames.

Besides the visual camera information there are other sensor types (e.g., temperature indications, CO levels) with valuable information for the incident commander. Visualizing too many sensors will lead to enacted sensemaking where an individual will select unconsciously some parts of the data to make their decision. Especially in critical or stress situations this can lead to overshooting or undershooting of the situation. To solve these issues the fire fighting community should get aware of the opportunities of web dashboards to quickly and easily create reports and data visualizations [2]. The core idea of dynamic dashboards is to efficiently display data in such a way that it can enhance user insight into the data. An example is the fire probability calculation and visualization for residential regions by Netage[1]. Similarly, the Moses project[2] explored a dashboard system for seamless localization, employability and health status monitoring (e.g., heart rate, air consumption) of firefighters. In Section 5.2.2 an investigation is done into the sensor/dashboard needs for fire incident management.

## 5.2 Visible and cognitive disability

Biologically, human beings have psychological restrictions in terms of observing and processing information. Furthermore, an individual can only process a limited amount of simultaneous stimuli. Ungerer et al. [3] stated that a person can only process a fraction of the information in a conscious way. According to his theory, each second 3 up to 5 visual and 3 auditive or tactile inputs can be processed simultaneously. Schaub et al. [4] on the other hand denoted that a human is only capable of processing 7 concurrent signals.

The head of operations and firefighters in general have to work in difficult, often unknown contexts, where there is an increased stress level. Subsequently, due to the increased level of stress there is a higher chance of reduced cognitive ability. Despite the dynamic and difficult character, decisions need to be made without detailed investigation or analysis. In general the fire scene is a complex situation due to the following factors:

- A comprehensive list of 'unknown' variables (e.g., number of people present, evacuation necessity, fire behavior, fire expansion and building contents;
- Variables that affect each other (e.g., ventilation openings will change the fire growth);
- Variables that change in time and space without being transparent or predictable.

Due to the risk of reduced cognitive ability and the psychological restrictions there is a need for a smart, adaptive visualization application. The time first-responders spend on gathering static and volatile situational information from affected people as well as from responsible persons at the emergency site can be reduced if they have access to the needed information. The following subsections will discuss the related work for fire incident visualization, the subjective criteria for data importance and some guidelines for usability testing.

[1] `https://netage.nl/`

[2] `http://techforfuture.nl/onderzoek/afgerond/mobile-sensing-safety-moses/`

### 5.2.1 First responders information needs: related work

Nunavath et al. [5] identified the information needs for first responders from a literature review, fire drills and interviews. Prioritizing of information items is needed for different stages of building fire emergency response operations and to ensure a maximum of 7 concurrent decision signals. The most commonly mentioned information item in the work of Nunavath et al. is the building related information, which includes important information about the building layout plans, hazardous material location, resources location, floors, and rooms. Furthermore, the other most commonly mentioned information item was fire related information such as color, location and condition of the fire. The generation of these features in our framework has already been introduced in Chapter 3 and 4 respectively.

Li et al. [6] investigated the information sources used in current practice (year 2013) and those desired to be used in the future. Furthermore, Li et al. considered the information items needed by first responders and the availability of technological solutions to obtain them in the United Stated. The desired information resources that were reported in these tests are listed in Table 5.1. Furthermore, Li et al. mention that there are three timings where data should be delivered: before arrival, when arriving at the emergency scene and during the attack and the mitigation. In an automatic framework these three states should also be considered and evaluated. In order to obtain more recent and related variables for Belgium, a questionnaire was launched. The results are discussed in Section 5.2.2.

**Table 5.1** Desired information needs during fire incidents evaluated in LA.

| Information sources | % of total response |
|---|---|
| Emergency pre-plans | 94.5 |
| Fire system | 85.2 |
| Responder reports and communications | 80.5 |
| Building 3D models | 80.3 |
| Human memory and experience | 79.9 |
| On-site observations and descriptions | 76.6 |
| Positioning and navigation equipment | 76.6 |
| HVAC system | 58.6 |
| Elevator system | 50.8 |
| Security and access system | 35.2 |
| Occupancy schedules | 32.8 |
| Power system | 31.3 |
| Lighting system | 26.6 |

Hamins et al. [7] created the smart-firefighting roadmap and their main suggestions are listed below.

- Use of sensors on the fire ground to know the staffs location and to assist in situational awareness;
- Increased collection and utilization of data before the incident to aid in effective use of personnel and equipment;
- Enhance interoperability between data systems;
- Develop intelligent systems to assist with decision making.

Although, the roadmap gives some directions for further research, currently only limited implementations are available. This chapter proposes some technological blocks to start with 'smart-firefighting'. Finally, emergency responders have different roles to perform, different tasks to handle, and different modes to communicate, which lead to an insufficient view of the complete emergency situation at the emergency site. In that respect Nunavath et al. [8] presented a conceptual domain model. The model consisted of four components: event component, actor component, objective component and building component. Each component contains several information resources and all components capture the complete building fire emergency response. However, as stated in Section 5.2 it is not feasible for a person to process and analyze all sensor components sequentially and filtering and summarization mechanisms will be necessary.

### 5.2.2 Subjective criteria and discussion

Video and sensor data are not processed and interpreted the same way by everyone, as indicated in Section 5.2. In that sense it is important to know the user's needs and wishes. Many times the 'importance scoring' is done by the IT designers or developers and this makes the application or dashboard not useful. Within the Belgian fire community, a survey was launched by HOWEST[3] and IDLAB[4] to explore the necessity and the use of new types of sensor data during a fire incident. This survey can be seen as an updated version of to the work of Li et al. [6]. Furthermore, an evaluation of different visualization set-ups was done with mock-up screens to gain further insights in fire data reporting.

The questionnaire was sent to all Flemish fire fighting communities and input was asked from all hierarchical levels (i.e., strategic, tactical and operational). Furthermore, we targeted a balance in the ratio of volunteer and professional firefighters that filled in the inquiry. 21 different questions were asked and the first question

[3] https://www.howest.be/nl

[4] http://idlab.technology/

**Table 5.2** Desired information needs during fire incidents evaluated in Belgium.

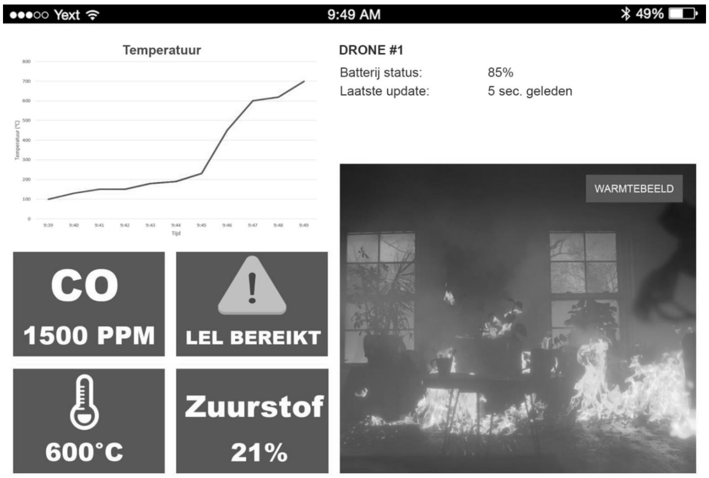
| **Information** | **% of total response** |
|---|---|
| Thermal camera | 95.9 |
| Explosion warning | 65.8 |
| CCTV | 65.8 |
| Localisation | 53.4 |
| Surrounding temperature | 47.9 |
| Percentage CH4 | 30.1 |
| Toxicity of the smoke | 28.8 |
| Chemical species | 28.8 |
| Ionizing substances | 27.4 |
| Percentage O2 | 21.9 |
| Percentage CO | 23.3 |

was: ”What kind of data should be available during an incident?”. From the survey results (see Table 5.2), it was remarked that the thermal imaging device is the most informative device used by firefighters, followed by the explosion warning device and the visual camera. In the questionnaire is was pointed out that the more information available during an incident, the higher the fire crew score their situational awareness. However, as stated earlier in this section, there is a maximum of 7 simultaneous inputs that can be used. The most important features are the location of the fire source, the amount and the position of the victims and firefighters, and the structure and lay-out of the building. All of them can be retrieved from thermal or visual cameras, as stated in previous chapters. Subsequently, in the questionnaire, some questions were related to the priority of the data. The camera footages, the thermal camera footages and the surrounding temperature were pointed out as prior. It is important to remark that this should be re-evaluated within a couple of years as most of the other sensor and data values are currently not available during an intervention and the decision makers are not trained to use certain variables. Furthermore, the questionnaires indicated that they allow a maximum set-up time (i.e., delay to connect to all devices) for the video streaming of 1 minute.

As important as the collection of the information is the visualization of the sensor data. The survey revealed that 78 percent prefers to use a rugged tablet, 20 percent selects a laptop whereas only 2 percent would use a smartphone device. Furthermore, depending on the incident type, different real-time data graphs need to be shown to facilitate the fire ground understanding. Furthermore, the respondents indicated that there should be a coupling with the existing data sources, such as the hydrant network or the statical intervention plans (i.e., a plan, only for large facilities, that contains contact details, the global outline of the building and the locations of dangerous goods).

To understand the end-user needs, their interaction and their interpretation, it is illustrated in literature that it is better to use an iterative design (i.e., where you continuously interact with the end-user). Several mock-up-screens are evaluated in the questionnaire and the most selected option was a visualization without tabs and with clear indications of excessive or abnormal values. Furthermore, the participants indicated the visualization of the most prominent videostream of the fire scene as an added value. Figure 5.1 and Figure 5.2 gives the two most selected visualizations.

**Figure 5.1** Simple visualization with live camera-stream and clear indication of dangerous values.




Mock-up displays are used to get initial feedback, but more usability testing will be necessary to have a stable, interactive and user-friendly system. In that respect, the following subsection will give more guidelines for the testing procedure.

**Figure 5.2** Visualization with two live camera-streams and graphs for temperature and oxygen level values.


### 5.2.3 Guidelines for usability testing

Technological excellence is not the only factor that drives market success of an application. Innovations should also tap into the end-users' unfulfilled needs. This subsection will indicate the evaluation mechanisms necessary to evaluate the user-driven fire incident dashboard that will combine the building blocks proposed in this thesis. Since the late 1990s, more and more attention has gone to measuring and understanding how users experience ICT products, applications and services. One of the drivers in this respect is that user experience (UX) has been linked to market success or failure of new and existing products, applications and services.

To investigate the end-user experiences, a multi-method, qualitative and quantitative mix of end-user research methods, customer-led idea generation and co-creation, should be set up [9]. One of the characteristics of a certain Living Lab methodology is that it gathers insights from the potential customer in a real-life context. Subsequently, the captured feedback is more reliable. This was shown by us in the SPOTT application [10].

After the application is developed and thoroughly tested on usability in a closed-prototype setting, the application should be tested in an open field environment. Furthermore a survey should be launched to focus on the usability of the fire incident management application measured by means of the System Usability Scale (SUS) [11]. The following paragraph gives some suggestions on how to interpret the SUS results. When a median SUS score is higher than 68/100, the usability is good. In the survey the potential for usage (i.e., adoption potential for current non-users) and potential for repeated usage (for current users) should be measured by means of the Product Specific Adoption Potential (PSAP) methodology proposed by De Marez et al. [12]. With the PSAP methodology, respondents are assigned to Rogers' innovation profiles (i.e., innovators, early adopters, early majority, late majority and laggards) [13] based on their answers to 3 statements (5-point Likert-scales: (1) *definitely use* and (5) *definitely not use*). The first statement measures the degree to which the participant would use the application. The second, optimal, statement measures the degree to which the participant would use the application with specific features. The third, sub-optimal, statement measures the degree to which the participant would use the application without the specific features.

## 5.3 Video summarization

As indicated in the questionnaire, decision makers want to have access to visual and thermal video streams. As it is not possible to analyze the complete history of a CCTV camera and it is not possible to investigate more then 7 cameras simultaneously, there is a need for video summarization. Video summarization can be classified into two types, namely static video summarization and dynamic video summarization. Static summarization results in a set of keyframes that best conveys the overall idea of the video. In most cases, keyframes are representative frames that could identify either the beginning or end of a scene transition in a video sequence. The number of resulting keyframes can vary as per the setting. Contrary, dynamic summarization collects small chosen fragments of the original video and arranges these fragments to obtain a new version of the video (i.e., logical story units).

Depending on the use case (e.g., fire incident, hazard materials incident) and the commanders function (e.g., evaluating the evacuation, investigating the structural damage or making strategic decisions for the intervention crews) different static or dynamic fragments will be indicated as important. Furthermore, personalized interaction will be necessary to learn the visualization needs based on the incident type, the operational function and the geographical location. The main focus of the following sections is on the static summary generation as it is not feasible to analyze several logical story units sequentially in a fast manner.

In order to visualize the relevant keyframes in a dashboard system and to decrease the computational cost of subsequent video processing tasks it is important to reduce the amount of video data by filtering out redundant and unnecessary frames, while preserving only those frames, distinctive and essential, to capture the entire video content. Furthermore, presenting the end-user with a limited list of representative keyframes instead of dynamic video fragments, improves their exploration, understanding and search process (see Section 5.2). Therefore the main focus of the following sections is on the static summary generation. The automated summarization of video content into representative keyframes, however, is a challenging problem due to the rapid change in lightning, viewpoint, and scene. In order to cope with these issues, we present a novel solution to extract, cluster, and filter meaningful keyframes. This summarization process consists of three major steps:

- The video summarization converts the video into a set of representative keyframes. All the keyframes represent one single video shot. These shots are detected using a grid-based histogram detection method and within each shot no-reference keyframe quality analysis is used to select the best frame within the shot.
- The number of keyframes is reduced by clustering visually similar frames, while preserving as much as possible of the entire content.
- A limited set of representative keyframes is given to the object and scene retrieval algorithm (see Chapter 3).

Much research has been done in the area of video summarization (i.e., keyframe retrieval). Ajmal et al. [14] give an overview of the different techniques and classification methods that are commonly discussed in literature, i.e., feature classification [15], clustering [16], shot selection [17] and trajectory analysis [18]. In the following subsection the grid-based histogram shot detection approach is proposed.

### 5.3.1 Shot detection

Compared to the state-of-the-art temporal shot segmentation algorithms, such as pixel/edge/motion differences based methods [19, 20], feature-based detection [21], spatio-temporal video slicing [22] and global histogram analysis [23–25], the proposed local histogram analysis on a 5-by-5 grid copes better with fast camera movements, zoom gestures, and similar scene discrimination (i.e., problems that arise in handheld video analysis). Object evaluation, based on manually generated ground truth data consisting of 400 video shots on different kinds of video content, results in a recall of 99% and a precision of 89.9% for our computationally efficient and fully automated process.

The proposed algorithm consists of 4 major steps:

- A conversion to gray scale frames.
- A rasterization (splitting) of each frame in a 5-by-5 grid and a calculation of the histogram for each cell in the grid.
- A correlation based temporal analysis of each histogram.
- A calculation of the amount of changed cells. If this is larger than an experimentally defined threshold the transition is seen as a shot.

The histogram based shot detection has an outstanding performance on detecting hard and gradual shots, which makes it suitable to process all kinds of video content (i.e., handheld and CCTV data streams). The only limitation is that it does not handle similar scenes very well, but this is solved in Section 5.3.3. In order to cope with gradual shots, like blends and fades, we also count the amount of frames that have been passed since the last detected transition. If that amount is too low and a new transition is found, we consider it to be the same transition. This way we successfully manage to detect both gradual- and abrupt scene transitions.

### 5.3.2 No-reference keyframe quality analysis

Currently, multiple image quality metrics are available [26]. However, most of them need a reference image or they are not suitable for real-time quality measurement. In our methodology, based on the detected shot boundaries, the frame with the highest quality within each shot is chosen as a representative keyframe. Subsequently, the quality is estimated using a weighted set of three no-reference quality metrics that are suitable for real-time images. Furthermore, the proposed no-reference exposure, contrast and sharpness metrics are evaluated on a variety of video footages. Evaluation results show that a combination of the computationally efficient metrics are sufficient in selecting the best quality frame(s) within each shot.

The no-reference exposure metric (Equation 5.1) results in a value ranging from 0 to 1. A value of zero means that the image is badly exposed, and the closer the value is going to 1 the more it is correctly exposed.

$$exposure = \frac{\bar{x} - 127}{127}, \tag{5.1}$$

*where $\bar{x}$ represents the average value of the histogram of the frame.*

The contrast (Equation 5.2) on the other hand is based on the normalized spread of the histogram of the frame. This results in a value of 0 for images without contrast and 1 for keyframes with a maximal amount of contrast.

$$contrast = \frac{1}{128}\sqrt{\sum_{i=1}^{N}(x_i-\mu)^2} \qquad and \qquad \mu = \frac{1}{N}\sum_{i=1}^{N}x_i, \tag{5.2}$$

*where $N$ represents the total amount of pixels in the frame and $x_i$ represents the i-th pixel value of the frame.*

Finally the sharpness (Equation 5.3) results in a value between -1 and 1, and is based on the fact that neutral lightning results in a histogram with all values centered around the middle of the keyframe histogram.

$$sharpness = \frac{1}{2*255^2*Width*Height}\sum_{i=1}^{Width}\sum_{j=1}^{Height}[(\frac{\delta G(i,j)}{\delta i})^2+(\frac{\delta G(i,j)}{\delta j})^2], \tag{5.3}$$

*where $G(i,j)$ represents the pixel value of the frame.*

### 5.3.3 Similarity clustering

In general, a video scene consists of many shots that are visually similar. Subsequently, an entire video sequence can contain several temporally distant yet visually similar frames that are taken in the same scene setting. On the one hand, removing these redundant and unnecessary frames improves the summarization visualization on a small screen or on a tablet. On the other hand, clustering the similar keyframes will decrease the time for object tagging and scene understanding (see Chapter 3), i.e., an automatic annotation can be performed on all object representations within the same cluster.

The clustering process is done in three steps. First, we perform a global feature extraction with CNN based learned features. Subsequently, we do a feature reduction by using the principal component analysis (PCA) technique. Finally, we use k-means clustering with the L2-distance (see equation 5.4) of the reduced features of the keyframe.

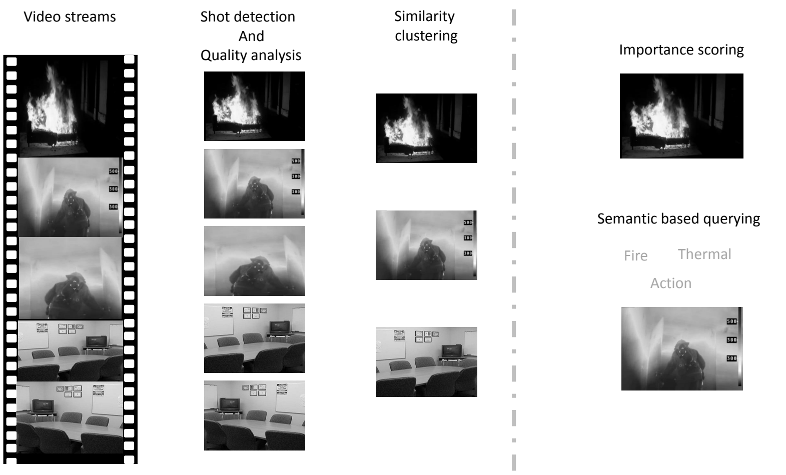
$$\sqrt{\sum_{n}^{i=1}(Vect1_i - Vect2_i)^2} \tag{5.4}$$

The learning-based features are generated by using a modified version of the Alexnet architecture [27] in which we removed the last three fully connected layers. The output after the last pooling layer is taken as the keyframe feature, as proposed by Zagoruyko et al. [28]. This ensures a higher level of image understanding and generalization and avoids high visual similarity scores just based on global color or texture information. Furthermore, it is important to remark that if the amount of clusters is too high, redundant keyframes will appear. On the other hand, if the amount of clusters is too low, outliers (i.e., very unique keyframes or single scene shots) will not be shown.

## 5.4 Video and frame retrieval

The set of keyframes (from different camera viewpoints and camera sources) could still be too versatile and to improve the fast exploration two additional mechanisms are proposed: (I) an automated importance scoring based on the scene content and (II) a semantic web-based querying tool. Furthermore, the tags that are generated from the scene understanding process (see Chapter 3, e.g., indoor, outdoor, person, closet, smoke, fire) are used as input for both mechanisms. Figure 5.3 gives a schematic overview of all the video summarization (see previous section) and retrieval blocks.

**Figure 5.3** Video summarization pipeline (left) and frame retrieval mechanisms (right)




### 5.4.1 Importance scoring

Based on the tags, as proposed in Chapter 3, we have information regarding the location, objects and people in every keyframe. The content of the scene is important because it includes useful information that attracts the viewer (e.g., the decision maker). Rationally, the objects and locations that are shown for a longer period of time in the handheld video have a higher chance to be of importance. It is important to remark that further user-driven evaluations will be necessary to prove this statement. Furthermore, since there are tags pertaining to all the frames it is necessary to alter them and to preserve only the most informative words that define the main semantic components. This is done as follows. First the tags are preprocessed by tokenization and part-of-speech tagging. After the extraction, the saliency score for each subset s is calculated according to the words they contain. We use Term Frequency - Inverse Document Frequency (tf-idf) [29] to weight each word (tag) $w$ as

$$tfidf(w,i) = \frac{(k_1+1)n_{i,w}}{k_1[(1-k_2)+k_2\frac{L_i}{A_L}]+n_{i,w}} \cdot log\frac{N}{n_w} \tag{5.5}$$

where $n_{i,w}$ is the number of occurrences of word $w$ in a set of keyframes $i$, $L_i$ is the number of words in the set of keyframes $i$, $A_L$ is the average number of words per set of keyframes in the video sequence, $k_1$ and $k_2$ are tuning parameters, $N$ is the total number of segments in the video sequence, and $n_w$ is the number of segments that contain word $w$. The tuning parameter $k_1$ determines the sensitivity of the first term $\frac{(k_1+1)n_{i,w}}{k_1[(1-k_2)+k_2\frac{L_i}{AL}]+n_{i,w}}$ to changes in the value of the term frequency $n_{i,w}$. If $k_1$ = 0, this term reduces to the counting function which is 1 if word $w$ occurs in segment and 0 otherwise. If $k_1$ is large, this term becomes nearly linear in $n_i, w$. Because not all the segments have same words, normalization is needed. $k_2$ is used to control the normalization degree. If $k_2$ = 0, there is no normalization. Following Blei et al [29], The tuning parameters are set as $k_1$ = 2 and $k_2$ = 0.75, but further user-driven sensitivity analysis will be necessary.

With the saliency scores for each individual word, the saliency score for a complete video sequence is calculated as the sum of the scores for the words it contains as in Equation 5.6.

$$T(i) = \sum_{w\epsilon i} tfidf(w,i) \tag{5.6}$$

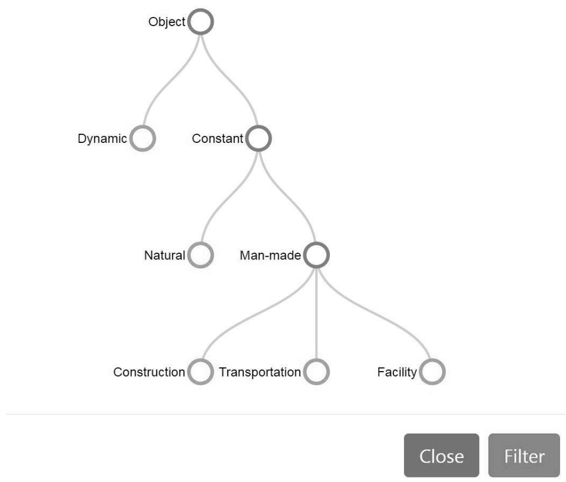
In combination with the similarity score (e.g., two visualized keyframes should be sufficiently dissimilar), the importance scoring is used to automatically present a couple of keyframes to the decision makers. Still, user evaluations and annotations could help improve and validate the automated video selection. Besides the automated selection of the keyframes, we propose a semantic querying system that allows to manually find the most related frames.

### 5.4.2 Semantic based querying

The web-based tool that could be used to explore the enriched (i.e., the annotated) keyframes is a metadata filtering and clustering service. The querying mechanism can be activated in three different ways:

- A textual input where the user can type the tag that he wants. In case the input is not in the list of predefined tags, the tag with the closest distance (i.e., the highest similarity) according to Wordnet [30] is selected.
- A drop-down list with predefined tags, where the user can select the necessary tags (i.e., show all the keyframes annotated with kitchen and fire).
- An interactive, hierarchical ontology visualization (see Figure 5.4) where the user can select and easily search for the best matching tags. Furthermore, if a tag on a higher semantic level in the hierarchy is selected, all the underlaying and related tags are selected (e.g., if the tag *opening* is selected, the tags *door* and *window* are likewise selected).

**Figure 5.4** Hierarchical, ontology driven interaction and visualization for keyframe tag filtering.




*An ontology is a representation, formal naming, and definition of the categories, properties, and the relations between the concepts, data, and entities over one or more domains*

The semantic similarity is currently calculated based on the ontology of Zhang et al., but new ontologies can easily be integrated. Zhang et al. [31] proposed a street scene ontology for qualitative understanding of outdoor scenes. This is valuable for large multi-disciplinary fire incidents as it contains building elements (e.g., floor, window, wall, column), but also construction, land and terrain elements. Some alternative ontologies that can be used are, for example, the work of Kadar et al. [32] who created a database of 100 scene categories (e.g., classroom, bathroom, bedroom, alley) derived from human vision. This ontology could be valuable in indoor fire fighting scenarios. Jaoa et al. [33] created an ontology on how scene situations progress in time. The FIRE ontology[5] was created in order to represent the set of concepts about the fire occurring in natural vegetation, its characteristics, causes and effects. Similar concepts and effects are found in indoor fire situations. Poveda et al. [34] created an ontology for designing and validating emergency plans and the sensor, users and furniture connections are highly valuable for our framework.

As indicated earlier in this section, the video summarization and retrieval building blocks require further user-driven evaluations. This, in combination with the new emerging technologies on technological or fire research, gives a great opportunity for more sensor and data-driven fire management.

Finally, the global work-flow is as follows: video footages from multidisciplinary incidents are taken as input for our evaluation. Figure 5.5 gives a schematic overview. First the video footages are selected (currently, this is a manual task, but the integration of online IP-cameras should be easily possible). Secondly, the keyframe extraction, similarity removal and no-reference analysis is used to select the most representable keyframes. Thirdly, the semantic tag understanding process is elaborated to automatically generate tags for each frame. Fourthly, the exploration and the retrieval of specific frames and situations is facilitated to get a fast overview of the current state of the incident.

[5] `https://bioportal.bioontology.org/ontologies/FIRE`

**Figure 5.5** Schematic overview of the video footage analysis framework, first the footage linking, secondly the keyframe generation, thirdly the content understanding tool and finally the exportability and retrieval tool for video investigation


## 5.5 Conclusions and future work

This chapter presented the user-needs and data restrictions specific for fire incident management. The insights were gained through the analysis of a questionnaire launched in the firefighting community of Belgium. This evaluation revealed that people are visually-oriented and that video footages are great to gain insights into a problem. Still, people can only process 7 image inputs simultaneous and for that reason, the video summarization framework, consisting of shot detection, frame quality and similarity analysis was proposed. Subsequently, in order to facilitate the video search process, the video and frame retrieval mechanism was clarified and semantic tag based querying on an existing ontology map was initiated. It is important to remark that the proposed techniques are equally suitable for commercial (e.g., soap series, cooking-shows) video summarization. Future work should investigate scalable and dynamic dashboards for multi-sensor input [35]. These dashboards should help with the automated indication of sensor and data needs during a fire incident.

# References


[1] Marvin M Chun. *Contextual cueing of visual attention*. Trends in cognitive sciences, 4(5):170–178, 2000.

[2] Olivier Janssens. *Data-driven performance monitoring, fault detection and dynamic dashboards for offshore wind farms*. PhD thesis, 2017.

[3] Friedrich Ungerer and Hans-Jorg Schmid. *An introduction to cognitive linguistics*. Routledge, 2013.

[4] Dietrich Drner and Harald Schaub. *Errors in Planning and Decision-making and the Nature of Human Information Processing*. Applied Psychology, 43(4):433–453.

[5] Vimala Nunavath, Andreas Prinz, and Tina Comes. *Identifying first responders information needs: supporting search and rescue operations for fire emergency response*. International Journal of Information Systems for Crisis Response and Management (IJISCRAM), 8(1):25–46, 2016.

[6] Nan Li, Zheng Yang, Ali Ghahramani, Burcin Becerik-Gerber, and Lucio Soibelman. *Situational awareness for supporting building fire emergency response: Information needs, information sources, and implementation requirements*. Fire safety journal, 63:17–28, 2014.

[7] Anthony P Hamins, Nelson P Bryner, Albert W Jones, Galen H Koepke, Anand Raghunathan, et al. *Smart firefighting workshop summary report*. Technical report, 2014.

[8] Vimala Nunavath, Andreas Prinz, Tina Comes, and Jaziar Radianti. *Representing fire emergency response knowledge through a domain modelling approach*. 2016.

[9] Dimitri Schuurman. *Bridging the gap between Open and User Innovation?: exploring the value of Living Labs as a means to structure user contribution and manage distributed innovation*. PhD thesis, Ghent University, 2015.

[10] Florian Vandecasteele, Karel Vandenbroucke, Dimitri Schuurman, and Steven Verstockt. *Spott: On-the-Spot e-Commerce for Television Using Deep Learning-Based Video Analysis Techniques*. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 13(3s):38, 2017.

[11] John Brooke. *SUS: a retrospective*. Journal of usability studies, 8(2):29–40, 2013.

[12] Lieven De Marez and Gino Verleye. *Innovation diffusion: The need for more accurate consumer insight. Illustration of the PSAP scale as a segmentation instrument*. Journal of Targeting, Measurement and Analysis for Marketing, 13(1):32–49, 2004.

[13] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[14] Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abbas, and Faiz Ali Shah. *Video summarization: techniques and classification*. In Computer Vision and Graphics, pages 1–13. Springer, 2012.

[15] Feng Wang and Chong-Wah Ngo. *Summarizing rushes videos by motion, object, and event understanding*. IEEE Transactions on Multimedia, pages 76–87, 2012.

[16] Luciana dos Santos Belo, Carlos Antônio Caetano, Zenilton Kleber Gonçalves do Patrocínio, and Silvio Jamil Ferzoli Guimarães. *Summarizing video sequence using a graph-based hierarchical approach*. Neurocomputing, pages 1001–1016, 2016.

[17] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. *Shot and scene detection via hierarchical clustering for re-using broadcast video*. In International Conference on Computer Analysis of Images and Patterns, pages 801–811. Springer, 2015.

[18] Xuekan Qiu, Shuqiang Jiang, Huiying Liu, Qingming Huang, and Longbing Cao. *Spatial-temporal attention analysis for home video*. In IEEE International Conference on Multimedia and Expo, pages 1517–1520. IEEE, 2008.

[19] S. Dagtas, W. Al-Khatib, A. Ghafoor, and R.L. Kashyap. *Models for motion-based video indexing and retrieval*. Transactions on Image Processing, IEEE, 9(1):88–101, 2000.

[20] Ramin Zabih, Justin Miller, and Kevin Mai. *A feature-based algorithm for detecting and classifying production effects*. Multimedia Systems, 7(2):119–128, 1999.

[21] Srinivasa Rao Chalamala, Krishna Kakkirala, and Jyoti Dhillon. *A robust video synchronization method based on hierarchical shot detection*. In International Conference on Audio, Language and Image Processing (ICALIP), pages 206–210. IEEE, 2014.

[22] A. Saoudi and H. Essafi. *Spatio-temporal video slice edges analysis for shot transition de-tection and classification*. International Journal of Control, Quantum and In-formation Engineering, 1(4):831–836, 2007.

[23] Won-Hee Kim and Jong-Nam Kim. *An adaptive shot change detection algorithm using an average of absolute difference histogram within extension sliding window*. In 13th International Symposium on Consumer Electronics (ISCE), pages 394–397. IEEE, 2009.

[24] R. Mishra, S. K. Singhai, and M. Sharma. *A Comparative based study of different video-shot boundary detection algorithms*. International Journal of Advanced Research in Computer Engineering & Technology, 2(1):282–289, 2013.

[25] Tian-Rui Liu and Shing-Chow Chan. *Automatic shot boundary detection algorithm using structure-aware histogram metric*. In 19th International Conference on Digital Signal Processing (DSP), pages 541–546. IEEE, 2014.

[26] K. R. Joy and E. G. Sarma. *Recent Developments in Image Quality Assessment Algorithms: a Review*. Journal of Theoretical & Applied Information Technology, 65(1):192 – 201, 2014.

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks*. In Advances in neural information processing systems, pages 1097–1105, 2012.

[28] Sergey Zagoruyko and Nikos Komodakis. *Learning to compare image patches via convolutional neural networks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4353–4361, 2015.

[29] David M Blei, Andrew Y Ng, and Michael I Jordan. *Latent dirichlet allocation*. Journal of machine Learning research, 3(Jan):993–1022, 2003.

[30] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. *WordNet:: Similarity: measuring the relatedness of concepts*. In Demonstration papers at HLT-NAACL 2004, pages 38–41. Association for Computational Linguistics, 2004.

[31] *Representing place locales using scene elements*. Computers, Environment and Urban Systems, 71:153 – 164, 2018.

[32] Ilan Kadar and Ohad Ben-Shahar. *SceneNet: A perceptual ontology for scene understanding*. In European Conference on Computer Vision, pages 385–400. Springer, 2014.

[33] Paulo A Joo, Joo Almeida, Patrcia Dockhorn Costa, and Giancarlo Guizzardi. *Towards an Ontology of Scenes and Situations*, 06 2018.

[34] Geovanny Poveda, Emilio Serrano, and Mercedes Garijo. *Designing emergency management services by ontology driven social simulation*. IT CoNvergence PRActice, 3(1):17–32, 2015.

[35] Olivier Janssens. *Data-driven performance monitoring, fault detection and dynamic dashboards for offshore wind farms*. PhD thesis, Ghent University, 2017.

# 6
# Conclusions and future work

In this dissertation we proposed the combination of state-of-the-art computer vision techniques and BIM for facilitating the localization and situational awareness problem in a fire context. The different building blocks that have been investigated, together with their interactions and their results, are shown in Figure 6.1. In the following sections, we list our contributions on each of these topics and we point out directions for future work. Subsequently, achievements and collaborations due to the research initiatives are explained in Section 6.4. Finally, an answer is given to the main research question initiated in Chapter 1.

## 6.1 Contributions

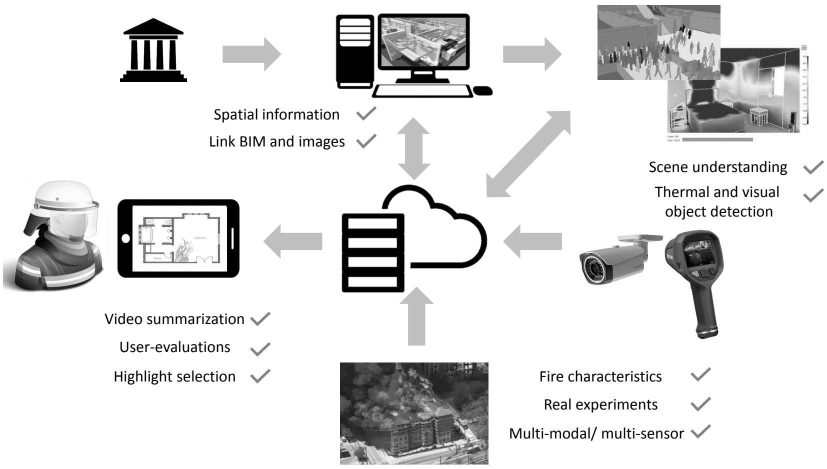
Investigation and exploration of spatial information (BIM and GIS) for fire safety design, evacuation and fire investigation was our first contribution. To the best of our knowledge it is the first approach that combines outdoor and indoor spatial information for fire science. The second novelty of Chapter 2 was the semantic matching of computer vision techniques and BIM data to facilitate localization and situational awareness problems in fire emergency situations. It is important to remark that building models need to be up-to-date, complete and rich in detail in case they are used for decision-making purposes. Currently, however, the building models are only used during construction and are mostly not updated afterwards, limiting their practical applicability.

In recent buildings there are alarms panels indicating the activated detection and suppression zones in case of a fire. However, this does not give information on the burning item or the detailed fire source location. Besides the smoke detection devices, there is an increased use of video footages, mainly for security and intrusion detection. The large number of videocameras should be exploited for fire incident management, for example, to give the status of the evacuation routes, to identify the amount of people inside the building, to estimate the smoke density or to define the fire fuel packages. In Chapter 3 a framework was proposed for automated fireground understanding from thermal and visual images and to increase the information flow. Subsequently, it was shown that a multi-labeling CNN is able to detect object and scene characteristics in a challenging environment. Furthermore, as it is necessary to build a system that can process the videostreams in real-time an optimization of the scene understanding framework was proposed by using motion detection, object tracking and model optimization. Finally, the contextual scene exploitation mechanism (i.e., the combination of object and scene detection probabilities) improved the visual and thermal scene understanding.

In Chapter 4 we initiated the generic architecture of the fireGIS framework, which allows the generation of real-time heatmaps that show the space-time distribution of fire risk levels and fire behavior characteristics. Furthermore, new tools were proposed for smoke visibility and flame height estimation. In order to show the feasibility of the proposed platform, real-fire experiments have been performed in a large-scale road tunnel and in a multi-compartment set-up. Subsequently, to ensure reliable sensor values at each spatial location in time a within- and between-variance analysis was suggested to improve the final accuracy. The proposed fireGIS framework extends the multi-modal/multi-sensor fire detection work that has been performed during the past years at IDLAB and Ghent University.

In the final chapter a clear explanation was given on the visible and cognitive disability in stress situations. People are visually-oriented and video footages and images are of great value to gain insights in a particular problem. Due to the fast decision making it is not feasible to analyze all video footages simultaneously. To solve this problem a video summarization mechanism and tools for highlight selection were proposed in Chapter 5. The novel video summarization framework combines shot detection, no-reference keyframe quality and similarity analysis. These mechanisms are suggested for fire incident video summarization, but they are equally suitable for commercial (e.g., soapseries, cooking-shows) video summarization. Finally, the subjective criteria and guidelines for usability testing are retrieved from a questionnaire launched in the Belgian firefighting community. User-driven evaluations are illustrated to increase the uptake of an application or device.

**Figure 6.1** Fire incident management interactions and realizations.




## 6.2 Future work

Future work directly in the line of this dissertation is evaluating the performance of the different algorithms and frameworks in more challenging and real fire environments. Furthermore, an automated collection of infrared images from real fires could increase the performance of the proposed scene and object detection mechanisms. This multidisciplinary research is a first step to make more reliable fire predictions. Still there are many opportunities to improve the multi-sensor fire behavior analysis. In that sense this dissertation can be seen as a road-map with building blocks for a more sensor-driven fire science.

### 6.2.1 The future of BIM for fire safety science

Due to the rising urban density, the aging infrastructure and the reduced budgets there is a need for a tool to enhance the current building lifecycle (i.e., planning, design, building and operation). BIM and GIS deliver such a tool to facilitate the data during the cycle. Since 2016, all centrally-procured construction projects are obliged to use BIM in the UK. In Spain, Germany and France there is an emphasis to use BIM for commercial and infrastructure projects by 2020[1].

[1] http://cupastone.com/bim-countries-world/

In general BIM is a relatively new technology and the potential benefits are not always that clear as the main focus is currently cost savings in the design and development phase. The BIM data can, for example, be used to schedule the most efficient maintenance or usage of a space. In this thesis it was already stated that the spatial data can help to reduce the human and economic loss due to a fire (see Chapter 2).

To make the usage of BIM data during a fire feasible there is need for an official government mandate on the storage and management of BIM data in Belgium. In that context a topic that has not yet been addressed in the thesis is the security objectives of the building model, which consists of three parts:

- **Integrity**: the first thing to consider is how to recover from failure or error conditions during the design phase. Several people with different backgrounds will work on the BIM model and interoperation problems can occur (e.g., the building owner changed the latest design and the fire modelers used an older version of the BIM file to calculate the expected fire evolution). Related to the fire science context it is also possible that malicious people 'infect' or modify the BIM model making the model useless during critical incident management. This repeats the need for indoor verification and validation mechanisms on a temporal basis (e.g., each year). Such a check-up could (for instance) be combined with the yearly assessment of the smoke detectors and the fire extinguish equipment.

- **Confidentiality**: control and authorization of the access rights to specific information or data. For the BIM usage in a fire safety science context it is quite clear that a terrorist or criminal should not have the same access rights compared to a building owner or a fire officer. A role-based access control allowing monitoring or accessing a specific part of the BIM and a specific level-of-detail is necessary. Finally, the access should be limited in time and not granted for the full lifecycle of the building.

- **Availability**: the building model should be unified accessible on a central (cloud-based) platform. Local, modified versions should be avoided. As, there is a higher chance that a fire or incident will occur in an older building than a more recent one, the compatibility of the software packages should be as high as possible. Furthermore, the software industry is evolving quickly and considerations should be made on how the BIM and fire science related data are stored and modified across the lifecycle (i.e., 50 years and more, from the earliest design up to demolition).

Based on the previous security objectives we suggest that there should be a centralized mechanism in the government that controls and facilitates the BIM files. Currently the files are stored in a decentralized manner making them not accessible during an incident. However, we expect that this will change in the upcoming years since more and more researchers start to explore the BIM-data and this in a wide range of application domains (e.g., augmented and virtual reality, IoT and open-data).

As stated in the Introduction, fire models require a detailed input of fuel packages, room configuration and ventilation conditions. Furthermore, accurate CFD calculations are slow (days up to weeks for one model) and require manual validation and verification. In that sense we suggest a future research track that investigates the feasibility of running fire models in advance for several ventilation and fuel conditions in the BIM file. In case of a fire, the best fire model is then estimated via an inversed modeling approach as proposed by Jahn et al. [1] and Rein et al. [2]. Still, the scalability of predicting fire spreading in real time will require more research and industry initiatives. One particular research question, for example, is how to access all direct (e.g., smoke detector) or indirect (e.g., presence detectors, HVAC temperature sensor) relevant sensors during a fire incident. Another research question is what level of simulation is required to facilitate a decent incident management.

### 6.2.2 The future of video fire analysis

Over the last years the cost of visual and infrared consumer devices has heavily decreased. This has led to an increase of commercially available video surveillance applications in a broad range of application domains (e.g., intruder detection and fire detection). Besides the fixed and PTZ (Pan, Tilt and Zoom) cameras, which are already available in many building environments, there are dynamically deployable cameras (i.e., a drone or a smartphone) that can give live feedback of the fire scene. Currently, the fire brigades are in an exploratory phase of the usage of certain dynamically deployable devices and more user-driven studies will be necessary. The guidelines proposed in Chapter 5 in this thesis can be used as a starting point.

More wireless devices, increased resolution, reduced camera size in combination with deep learning trends will define video analysis in the future. Furthermore, the performance of the analytics algorithms will only increase in uncontrolled situations. For example, only a couple of years ago face detection was only possible in controlled indoor settings. Now it is possible to estimate emotions and expressions of a face with a regular smartphone camera.

Future work on video fire analysis should investigate the material type recognition with standard visual and hyperspectral cameras. Objects are made of different materials and linings and different fire behavior can be expected. Each object has another reflectance and transmission influenced by thermal radiation received from other materials. Therefore, future work should explore the combination of different low-cost sensors (e.g., depth, visual and thermal) to improve the material recognition. In a similar way, the estimation of 3D indoor object characteristics, will only become easier in the following years due to the possibilities of deep learning (e.g., single sensor depth estimation, auto-encoders for 3D shape recognition) as stated in Chapter 3.

Wang et al. [3] stated that CNN models are even capable of identifying the HRR of a fire from a single video footage. However, as the fire growth is dependent on the maximal fuel package, the inlet and outlet ventilation conditions, the surroundings and the compartment configuration, more video footages in combination with building information models will be necessary to calculate the HRR. Multi-sensor inputs in combination with BIM data seem here the most appropriate direction as proposed in Chapter 2. Soon, video analytics will be able to perform person recognition, crowd analysis (e.g., detecting the amount of people and their behavior) and anomaly detection on global scale. Maybe in the future, even before a fire incident will occur the video analysis will notify suspicious behavior and faster interventions could reduce the economic and material damage.

## 6.3 Emerging technologies and trends

In this section we want to address some upcoming trends and technologies that are valuable for fire incident management. Drones, automated robots and handheld thermal cameras are emerging in different fields.

### 6.3.1 Drones for incident management

In the final part of the survey (see Section 5.2.2) we investigated the current usage of drones, also referred to as unmanned aircraft systems, for firefighting incidents. It is important to remark that, due to the limited flight time and the non-resistance to high temperatures, there are currently no systems available for indoor fire incident exploration. Still, due to their speed and outdoor reach it is already possible to assist in risk assessment, mapping, planning and reducing the exposure to danger of firefighters. Future research is necessary to create protected drones and to ensure a stable communication on the incident site, but preliminary results show the added value that drones can provide in fire incident management.

### 6.3.2 Handheld thermal imagers

The image quality of handheld thermal cameras is increasing while the price is decreasing. As a result, more and more fire brigades actively use these devices. Nevertheless, there are currently limited practice guides to use these imaging devices. Amon et al. [4] investigated the problems to use a thermal imaging device in a firefighting context, but with limited focus on the correct handling of the device. In order to fully integrate our proposed object detection mechanism for fireground understanding, more research on this topic is needed. Meanwhile, some practical advices for proper use are given based on our current expertise:

- Point the camera forward to get a complete overview of the room; usually during a fire incident the camera is only pointed upwards,
- Avoid the presence of persons (i.e., other firefighters) in the field-of-view to ensure a maximal detection and recognition of objects,
- Avoid the pointing on reflective surfaces, such as mirrors or metal doors, to avoid false positives for the object detection,
- Move the camera slowly to prevent the image from freezing and to avoid/limit the creation of artifacts,
- While moving into a building, occasionally look behind you to make sure the path you have taken is clear, and keep track of any smoke or fire changing conditions,
- It is important to know that a thermal camera cannot look through furniture or underneath debris. Flames or furniture objects cannot be detected if they are not in the field-of-view.

Besides the research on the application of thermal imaging devices in a firefighting context there is ongoing research to make the handling of a thermal device easier. Scott Sight built an in-mask thermal sensor[2] and VIZIR[3] offers a hands-free operation mask. There are some differences between both systems. For instance, the Scott Sight has no external camera and a thermal viewer in the corner, while the VIZIR has its augmented reality images in the center. Further research will be necessary to achieve full fireground understanding from these devices.

[2] http://www.scottsafetynation.com
[3] http://www.darix.ch/

In addition to the active monitoring of the fire fighters' performance it is important to get a clear view on their environment. In spite of the active use of thermal devices, however, there is currently no system available to perform automated fire location understanding. In this thesis we described some technological innovations that could help create an automated overview of the environment during intervention. Each of these technologies will contribute in transforming the firefighting into a more sensing and data-driven discipline.

Finally, another recent trend in the firefighting community is the use of robots, mostly for urban search. Still most of these robots are in an exploratory, research phase and require manual manipulation of the device. The TRADR project [5], for example, showed a robot-assisted response to explore the disaster scene. The SAFFIR project created a prototype robot to walk across uneven floors, use thermal imaging to identify overheated equipment or fire, and even hold a hose to extinguish a small fire.

## 6.4 Achievements and collaborations

The work initiated in this thesis will enable the semantic intelligence research unit of IDLAB to further extend activities in multimodal data processing, video analysis, computer vision, GEO-ICT, spatio-temporal enrichment and data mining/-analysis. Subsequently, the methodologies and tools proposed in this thesis will improve the spatio-temporal (meta) data quality and querying process.

Due to the insights gained in this thesis several new research initiatives were initiated. The scene and object understanding building blocks of Chapter 3 inspired the UGESCO[4] Belspo project. The goal of the project is to develop geo-temporal (meta)data extraction and enrichment tools to extend and link the existing digital archive collection items and facilitate spatio-temporal collection mapping for interactive querying.

The spatial information analysis of Chapter 2 and 4 contributes to the SmartGLAZ imec-ICON project. Spatio-temporal environmental information (e.g., landmarks, hot spots and weather data) will be used to create environment-aware, continuously self-configuring projection in a Head Mounted Display (HMD).

The work of video summarization (see Chapter 5) and object-scene relations (see Chapter 3) are contributions to the SPOTT IWT research project[5]. The goal is to enrich video content and to facilitate the keyframe retrieval.

[4] http://ugesco.be/
[5] https://www.imec-int.com/nl/istart-portfolio/spott

Finally, due to the initiatives that were taken during this research a new course will be organized in the Master Fire Safety Science in the University of Ghent from next academic year. The aim of the course is to provide students insights into the different building blocks of sensor and data-driven management of fire incidents.

## 6.5 Answer to the research question

In order to end this dissertation we reformulate the central question initiated in Chapter 1 and answer the different aspects:

> **Can we develop a system to accurately detect, analyze and visualize spatio-temporal fire characteristics in enclosure fires and can we use the extracted information for fire behavior analysis and forecasting within a BIM framework?**

The first part of the research question was **the linking with a BIM framework**. The literature review and suggestions in Chapter 2 of this thesis showed the feasibility of the integration of BIM data for fire investigation. As indicated earlier it is important to remark that building models need to be up-to-date, complete and rich in detail.

The second part of the central research question, i.e., **in enclosure fires**, required a good knowledge of the room configuration, the scene type and subsequently an understanding of enclosure fire dynamics. In Chapter 3 we have shown that it is possible to estimate in real-time the scene type and the indoor object configuration. Besides the object knowledge, additional semantic information of the scene classification module could be used to further improve the localization accuracy. Finally, the feasibility of our system was illustrated in a large-scale road tunnel and in a multi-compartment set-up.

The third aspect was the **accurate detection, analysis and visualization of spatio-temporal fire characteristics**. In Chapter 4 we proposed the generic architecture of the fireGIS framework, which allowed the generation of real-time heatmaps that show the space-time distribution of fire risk levels and fire behavior characteristics.

The fourth aspect of using **the extracted information for fire behavior analysis and forecasting** links to two facets. Firstly, the analysis links to the decision making and the situational awareness. In Chapter 5 user requirements and guidelines for fire incident management were proposed. Furthermore, the video summarization tools help to improve the situational awareness. Secondly, the forecasting assumes that the retrieved information can be incorporated into existing fire forecasting CFD or zone models. However, the problem for sensor based fire forecasting in real-life use-cases has not been solved yet. This work gives a major base for the extraction of smoke, building and fire characteristics and future work should focus on improving the incorporation of the sensor values in existing forecasting applications.

This thesis shows the technology readiness of each of the proposed building blocks. It should be seen as a starting point to make the fire fighting more sensor and data-driven. Furthermore, it is important to remark that several of the proposed building blocks are not limited to fire science and they can be easily adapted to other applications, e.g., multi-model object recognition in shopping content, commercial video summarization and querying or adaptive indoor routing. In general the results in the thesis have a broad scientific importance and the usability of our contributions will only increase in the coming years.

# References

[1] Wolfram Jahn. *Using suppression and detection devices to steer CFD fire forecast simulations*. Fire Safety Journal, 91:284 – 290, 2017.

[2] W Jahn, G Rein, and JL Torero. *Forecasting fire growth using an inverse zone modelling approach*. Fire Safety Journal, 46(3):81–88, 2011.

[3] Zhenyu Wang, Chunfeng Song, and Tao Chen. *Deep learning based monitoring of furnace combustion state and measurement of heat release rate*. Energy, 131:106 – 112, 2017.

[4] Francine K Amon, Nelson P Bryner, and Anthony Hamins. *Thermal imaging research needs for first responders: workshop proceedings*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2005.

[5] Ivana Kruijff-Korbayová, Francis Colas, Mario Gianni, Fiora Pirri, Joachim de Greeff, Koen Hindriks, Mark Neerincx, Petter Ögren, Tomáš Svoboda, and Rainer Worst. *Tradr project: Long-term human-robot teaming for robot assisted disaster response*. KI-Künstliche Intelligenz, 29(2):193–201, 2015.

# A
# Introduction to fire behavior

**Abstract** *This appendix provides an overview of basic fire behavior concepts and physics, to give some background and to guide the reader through this dissertation. Subsequently an introduction is given into enclosure fires and finally a preface is given into CFD and zone modeling for fire forecasting concept. Major parts of this chapter are inspired by the work of Merci et Beji. [1].*

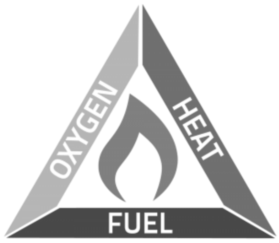
## A.1 Introduction

The first question: ”What is a fire?” should be answered before going more in detail into enclosure dynamics. A fire needs 3 elements (see Figure A.1): oxygen, fuel and a heat source. A candle flame clearly illustrates these three fire triangle values, (i.e., the fuel is the wax, the oxidizer is the oxygen in the air and the heat is initially brought by the ignition and sustained by the flame). Furthermore, a fire is characterized by heat and smoke. The smoke is mainly described by the visibility and the toxicity and the heat is defined by a temperature and a flux.

### A.1.1 Heat transfer in fires

Heat transfer is key in the context of fire and has been widely discussed in different textbooks (e.g., Drysdale [2] and Merci et Beji. [1]). The heat supplied to the fuel can transform the fuel from solid, up to liquid and gaseous phase. This is done by thee different mechanisms: conduction, convection and radiation.

**Figure A.1** Fire triangle representation (source: Elite Fire Protection Ltd).




- Conduction: is the transfer of heat within the material itself, the heat transfer occurs from a region of higher temperature to a region with a lower temperature. Specific for fire spreading, two examples can be given where conduction plays a role: the walls are heating up and the heat is transmitted to the adjacent walls and spaces; insulation: wood and gypsum specific, are poor conductors and they transmit the heat very slowly.

- Convection: transfer of heat by the physical movement of fluids (air or another carrier) in contact with a surface. When the air is heated, it expands and the air becomes lighter than the surrounding air and it rises. When a fire is burning large amounts of hot gases and smoke, one of the most important factors affecting life safety, are produced. These will travel through the building and often ignite more combustible materials causing the fire to spread.

- Radiation: the emission of energy in rays or waves without any mass transfer. The heat from the rays can be absorbed by combustible materials which can cause them to heat up and to ignite. For indoor fire spreading the radiation plays an important role when the thick, hot smoke layer is formed and spreads across the room. The hot layer will radiate the underlying objects and can induce the roll-over and flash-over.

### A.1.2 Definitions

**Pyrolysis:** refers to the degradation of solid material and is an irreversible endothermic process. In contrast to evaporation it changes the chemical composition of the material.

**Neutral layer:** The interaction plane between the cold layer and the hot smoke layer is in literature called the neutral plane or layer. This is not always clearly visible in reality due the turbulence of the fire and the smoke.

**Heat of combustion:** Drysdale [2] defined the heat of combustion as the total amount of energy released when a unit quantity of a fuel at 298 Kelvin and at atmospheric conditions is completely oxidized. This value will indicate the behavior of the material/fuel in case of a fire.

$$\Delta h_{c,eff} = \frac{\dot{q(t)}}{-\dot{m}} \tag{A.1}$$

$\Delta h_{c,eff}$ = effective heat of combustion [MJ / kg]
$\dot{q(t)}$ = heat release rate in time [kW]
$\dot{m}$ = mass loss rate of specimen [g . $s^{-1}$]

**The Heat Release Rate (HRR):** is the rate of heat generation by a fire and it is a critical parameter to characterize the fire. Furthermore, the parameter can be used to evaluate the fire growth of an enclosure fire and for fire assessment, hazard analysis or computational fluid dynamics. From the previous formula A.1, the heat release rate can be determined. Secondly, the heat release rate can be determined accordance with the ISO 5660 (second edition) international standard document, by measuring the oxygen consumption and the flow rate of the combustion products in a cone calorimeter. In Figure A.2 there is a comparison for the peak heat release rate for different piloted and radiated ignited ($35kW/m^2$) calorimeter fires in accordance to the ISO 5660 test, the tests are performed in WarringtonFire Ghent. The disperse material forms have a higher heat release rate compared to the massive form. This is due to the amount of oxygen that can be mixed with the fuel for burning (fuel bed area). Due to the ideal mixing environment, the burning is going faster, whereas massive materials have a more constant and longer burning when sustainable flaming.

## A.2 Fire growth

The growth of a fire (i.e., slow, medium, fast, ultra-fast) is indicated by the slope of the heat release rate (expressed in [kW]) as function of time. In literature [3] this is often represented by an $\alpha t^2-$curve.

Figure A.3 gives an overview of typical fire growth rates and corresponding indicative material types. As can be seen, a methanol pool fire will have an ultra-fast growth rate, compared to a solid wood cabin, which has a slow growth rate. The growth is mainly dependent on the room-configuration and content. Obviously, the flame spread rate affects the fire growth.

**Figure A.2** Peak Heat Release Rate for different materials in disperse and massive form

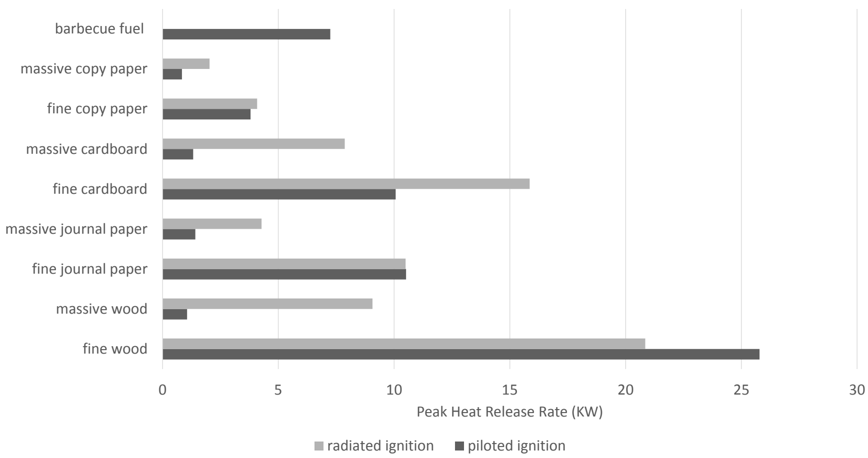
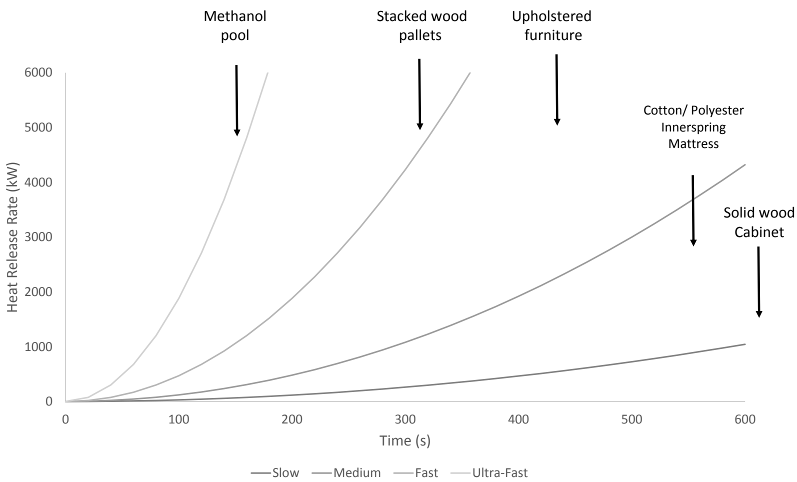



**Figure A.3** $\alpha t^2-$curve with different fire growth rates and indicative materials



### A.2.1 Flame spread

Flame spreading is in literature referred as a succession of ignitions that increasingly enlarge the affected area. Many factors affect the flame spread rate:

- The surface orientation and direction of propagation (e.g., vertical upward flame spread is very fast, due to strong radiative and convective heat transfer).
- The material thermal inertia (e.g., fire on light materials spread fast).
- Surface geometry (e.g., the flame spreading is faster in a corner fire compared to a similar sized fire in the middle of the wall)
- Environmental effects (e.g., increased fire spreading in a forest due imposed air movement).

To measure the flame height and eventually the fire spreading there are different methods proposed in literature. More detailed explanation of our proposed techniques and experiments is given in Chapter 4, but in literature two major mechanisms are mentioned.

- Based on the surface temperature, measured with thermocouples[1], the major flame front is indicated by a temperature above 300 °C [4].
- Based on a visual or thermal camera based flame detector algorithm [5, 6].

## A.3 Enclosure fire dynamics

When a fire occurs in an enclosed space, additional heat transfer and physics elements need to be considered [7]. First there is the ignition phase, secondly there is the growing phase, where several evolutions of the fire are possible. The fire can fade out or the fire can grow rapidly (due the positive feedback loop: the heat from the flames induces more pyrolysis of virgin material; more pyrolysis, flammable gases are released and larger flames appear. Subsequently, there is more heat-transfer to the non-burning objects and more pyrolysis). Thirdly, flash-over can occur (i.e., a rapid spread from the area of localized burning to all combustible surfaces within the room) and the fire becomes a fully-developed fire. Finally, the fire comes in a decay phase due to lack of fuel or oxygen. It is important to remark that the evolution in time and temperature is situation dependent. Furthermore, a sudden opening or breakage of a window can change the air supply rate and this can induce a ventilation-induced flashover.

[1] A thermocouple is an electrical device made of two dissimilar conductors. Due to the thermoelectric effect and the Seebeck effect (e.g., conversion of heat into electricity at the junction), different voltages can be measured at different temperatures.

**Figure A.4** Evolution in time of the indicative HRR in an enclosure within fuel controlled conditions.

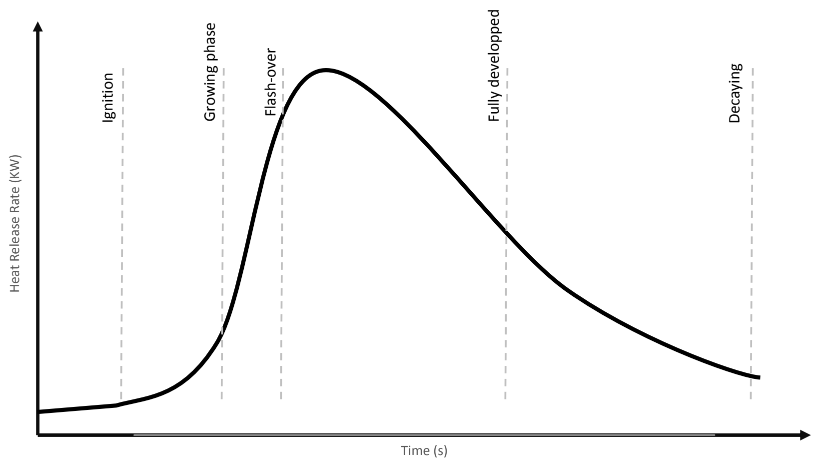



Figure A.4 shows the evolution in time of the indicative HRR in an enclosure with fuel controlled conditions. The flash-over state is from a fire safety science point-of-view a tipping point. The design for the pre-flashover state focuses on the safety of humans and the reaction to fire, whereas in the post-flashover state the emphasis is on the design of structural stability (and safety of fire fighters): fire resistance.

### A.3.1 Fuel versus ventilation controlled fire

In the growth phase the fire can be 'fuel controlled' or 'ventilation controlled'. In the fuel controlled case the fire can die if the heat provided by the flames is less than the energy required for further pyrolysis of virgin material, or the burning object is completely burned. If we look at the simplest combustion reaction of methane, in the fuel controlled case, there is more oxygen available than necessary. $CH_4 + 4O_2 - > CO_2 + 2H_2O + 2O_2$ For the ventilation-controlled case the fire can die if there is a lack of oxygen. Figure A.5 indicates the fuel controlled conditions of a compartment fire during the growth and the decaying phase.

**Figure A.5** Evolution in time of the indicative HRR for fuel and ventilation controlled situations (source cfbt-us.com).

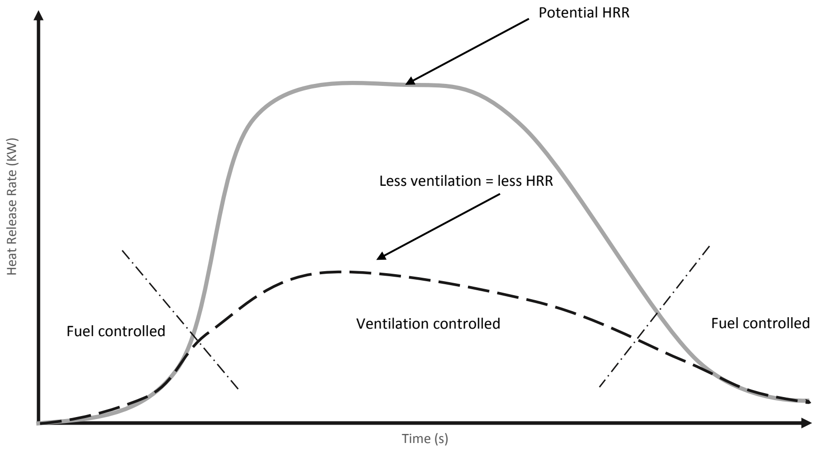
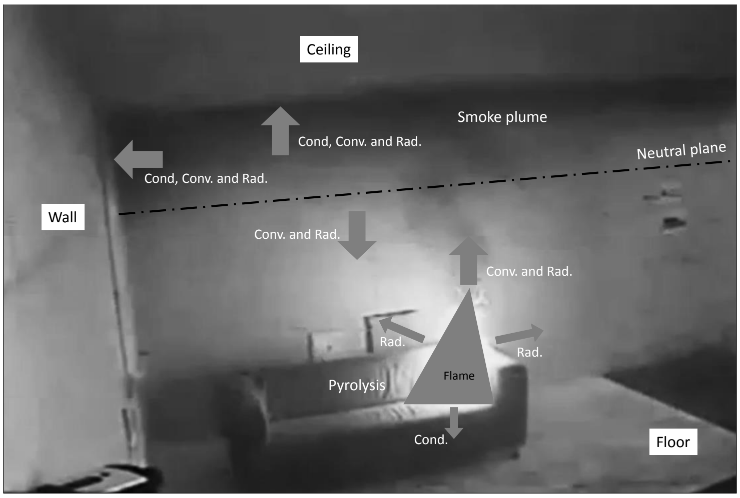



**Figure A.6** Simplified heat transfer in enclosures fires, couch fire during the VIPA-BVO experiment (source VIPA-study).




## A.3.2 Heat transfer in enclosure fires

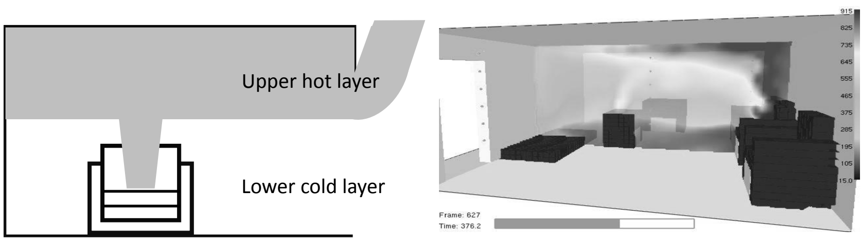
Besides the oxygen consumption, the fire stages and the heat release curves, it is also important to look at the heat transfer, which is not straightforward in enclose fires. Figure A.6 gives a simplified version of the different heat transfer interactions in an enclosure fire. The flame releases some heat by radiation to the smoke and to the surroundings. Subsequently, due the buoyancy (i.e., upward force), the heat is mainly transferred by convection into the smoke plume. Furthermore, due the radiation on the fuel surface, the pyrolysis process creates combustible gases. Another part of the heat on the fuel surface disappears by conduction and is responsible for the heating up of the solid material.

As indicated in Figure A.6 there is in the beginning a build-up of a smoke layer. This is possible due the bounding surface (i.e, ceiling, walls, floor). Furthermore, the layer heats up the walls by convection and radiation and the walls transport the heat internally by conduction. Finally, in later stages of the fire development, there is radiative feedback from the hot walls, ceiling and smoke layer, enhancing the pyrolysis and the rapid fire progress.

## A.4 Fire modeling

To understand the fire dynamics and ultimately to build fire-safe buildings, scientist and designers frequently use fire modeling tools. Based on the space geometry, the fuel package dimensions and the fuel properties an estimation of the fire representation is made. The modeling tools can, based on their complexity, be divided into three groups: algebraic models, zone-models and CFD models, where the latter one is the more sophisticated version. Within this thesis it is not the intention to optimize the existing methods and for more physical details reference is made to different textbooks [1, 8, 9]. Still to have a basic understanding a small introduction for each method is given. Finally, it is important to remark that all the models require expertise in defining the correct input data and assessing the feasibility of the calculated results.

**Figure A.7** Couch fire simulation: Left: the simplified zone model with only the cold and hot layer; right: the CFD model with detailed temperature profiles (source: M10 fire consultancy).



### A.4.1 Introduction to zone modeling

Zone models assume that the volume or room can be divided into zones with homogeneous temperature and species composition. In the enclosure fire case, there is the ideal hot layer with combustible products at the ceiling of the compartment and the cold layer with fresh air. The temperature and the height of each layer can change in time. This is a very strong simplification and the models can only be used in simple configurations (i.e., complex non rectangular geometry, stratification of hot gases). The computation time is only a matter of seconds and allows fire modelers to obtain fast results for different variables and configurations during the design phase.

### A.4.2 Introduction to CFD fire modeling

Computational Fluid Dynamics (CFD) is a tool to simulate, in this thesis fire-driven, flow through numerical solutions of the Navier-Stokes equations [10]. The solution is based on finite differences or finite volumes to discretize the partial differential equations, expressing the conservation of mass, momentum and energy. Compared to the zone-model, where there are two volumes, in CFD there are millions of volumes. The higher the number of volumes, the more accurate the final resolution. Contrarily, the computational calculation time and complexity increases (reducing the cell size by factor 2 will increase the computation time by factor 16). The major benefit compared to zone models is the capability of modeling smoke and heat movement in complex geometries in multi-compartments (i.e., atria and tunnels). Furthermore, due to the possibility to modify the room configuration, boundary conditions, addition of sprinklers and smoke detectors, the CFD models are often used to assess the influence of different parameters in the design phase of a building. Finally, there are many CFD packages available, but specific for fire behavior FDS[2] Firefoam[3] are most frequently used.

[2] https://pages.nist.gov/fds-smv/
[3] https://github.com/fireFoam-dev

# References


[1] Bart Merci and Tarek Beji. *Fluid mechanics aspects of fire and smoke dynamics in enclosures*. CRC Press, 2016.

[2] Dougal Drysdale. *An introduction to fire dynamics*. John Wiley & Sons, 2011.

[3] National Fire Protection Association et al. *NFPA 92B: Guide for Smoke Management Systems in Malls, Atria, and Large Areas*. National Fire Protection Association, 1995.

[4] Gunnar Heskestad. *Fire plumes, flame height, and air entrainment*. In SFPE handbook of fire protection engineering, pages 396–428. Springer, 2016.

[5] Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J Hollemeersch, and Rik Van de Walle. *FireCube: a multi-view localization framework for 3D fire analysis*. Fire Safety Journal, 46(5):262–275, 2011.

[6] A Enis Çetin, Kosmas Dimitropoulos, Benedict Gouverneur, Nikos Grammalidis, Osman Günay, Y Hakan Habibolu, B Uur Töreyin, and Steven Verstockt. *Video fire detection–Review*. Digital Signal Processing, 23(6):1827–1843, 2013.

[7] Lars-Göran Bengtsson. *Enclosure fires*. Swedish Rescue Services Agency, 2001.

[8] James G Quintiere. *Fundamentals of fire phenomena*. John Wiley Chichester, 2006.

[9] Morgan J Hurley, Daniel T Gottuk, John R Hall Jr, Kazunori Harada, Erica D Kuligowski, Milosh Puchovsky, John M Watts Jr, CHRISTOPHER J WIECZOREK, et al. *SFPE handbook of fire protection engineering*. Springer, 2015.

[10] Roger Temam. *Navier-Stokes equations: theory and numerical analysis*, volume 343. American Mathematical Soc., 2001.

# B

# Computer vision and machine learning basics

**Abstract** *This appendix provides an overview of basic computer vision and machine learning concepts and theories, to guide you trough this dissertation if you are not familiar with these techniques. The following sections are limited to the techniques used in this thesis. This appendix should therefore not be seen as a complete overview of machine learning and computer vision.*

## B.1 Computer vision

Before going more into detail in machine learning and deeplearning it is important to have some basic understanding of the most general image processing procedures. Furthermore this know-how can be used for preprocessing, image filtering and data augmentation. All of them are very important techniques to achieve high detection/recognition accuracy in real world applications, where noise, lightning changes, occlusions and reflections can occur in the video or images. Such images contain a discrete number of pixels and image processing takes place on that pixel level. Some examples of frequently used image processing are contrast enhancement, edge detection, noise removal and geometric transformations.

### B.1.1 Pixel representation

1. An image consist of a grid of pixels;
2. Every pixel has a coordinate (which describes the location in the grid, more specific the row and the column). Furthermore the pixel contains color and lighting information, this is represented by a set of numbers and the amount defines the depth of the image. A 24-bit image, for example, allows to store three 0-255 numbers for each pixel, for example, raw RGB values.

The RGB values represent a specific color space or model of an image. A color model is an abstract mathematical model that describes the way that colors will be represented as tuples of numbers. The RGB model, for example, will represent a color in a combination of primary color spectra (e.g., R = 700nm, G = 546,1 nm, B = 435,8 nm).

Two other color models that are used in this thesis (see Chapter 4) are the Hue, Saturation and Value (HSV) model and the YUV model. HSV gives an approximation of the human perception of light. Compared to the RGB-space, where the values are defined between 0 and 1 or 0 and 255, in the HSV-space the H is a value between 0 and 360, the S and V values are between 0 and 1. Figure B.1 gives an example of a car fire image where the upper left image is the original RGB representation and the other images are the transformed H, S and V values. Furthermore, Formula B.1 give the transformation from the RGB to the HSV space.

$$H = \begin{cases} \left(0 + \frac{G-B}{MAX-MIN}\right) * 60 & \text{if R = MAX} \\ \left(2 + \frac{B-R}{MAX-MIN}\right) * 60 & \text{if G = MAX} \\ \left(4 + \frac{R-G}{MAX-MIN}\right) * 60 & \text{if B = MAX} \end{cases} \tag{B.1}$$

$$S = \frac{MAX - MIN}{MAX}$$

$$V = MAX$$

The YUV spectrum, where Y stands for luminance (brightness) and the UV for the chrominance values, is also used in Chapter 4. The transformation from the RGB image to YUV is given in Formula B.2.

$$\begin{aligned} Y &= 0,299.R + 0,597.G + 0,114.B \\ U &= B - Y \\ V &= R - Y \end{aligned} \tag{B.2}$$

**Figure B.1** Image color space visualization, upper left the original RGB image, right the Y and H value, at the bottom respective the UV and the SV representation.

(a) YUV-space (b) HSV-space

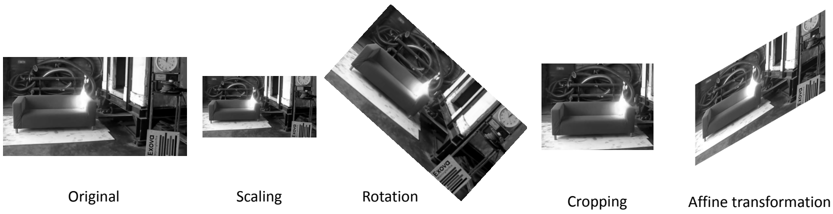
### B.1.2 Pixel-wise and geometric transformations

There are five main geometric transformations that can be performed on an image: scaling, translation, rotation, cropping and affine transformation. The operations do not change the image content, but deform the pixel grid and map it to the new destination image, see Formula B.3. These transformations are also the ones that are used in Chapter 3 to expand the scene and object dataset.

$$dst(x, y, z) = src(f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)) \tag{B.3}$$

**Scaling** is the resizing of the image with different interpolation methods (e.g., nearest-neighbor, bilinear, bicubic) along the coordinate directions. **Translation** is the shifting of the objects location in the x, y or z direction. **Rotation** provides scaled rotation with adjustable center of rotation and adjustable angles. **Cropping** is the selection of a region of interest from an image. **Affine transformation** is a basic transformation where all the parallel lines in the original image remain parallel in the output image. It does not always preserve the distance between the points in the original and destination image. However, it preserves the ratios of the distances between the points. Figure B.2 gives an overview of the different transformation mechanisms.

**Figure B.2** An overview of the transformation methods: scaling, rotation, cropping, affine transformation.




### B.1.3 Image smoothing or blurring

Image blurring is achieved by convolving the image with a low-pass filter kernel and it is useful for removing noise. It actually removes high frequency content. OpenCV[1], a well-known image processing software package provides mainly four types of blurring techniques. Averaging is the most general blur technique, Gaussian filtering is highly effective in removing gaussian noise from the image, Median filtering is highly effective against salt-and-pepper noise (sparsely occurring white and black pixels) in the images and bilateral filtering is highly effective in noise removal while preserving sharp edges. The latter technique (i.e., the bilateral filter) was chosen as a pre-processing step in the smoke visibility estimation algorithm proposed in this thesis.

### B.1.4 Morphological operations

Morphological operations apply a structuring element to an input image in order to improve object representations in the image. The applications of the operation are:

- Noise removal;
- Isolation of individual elements and joining disparate elements;
- Finding of intensity bumps or holes.

The most basic morphological operations are erosion and dilation. Figure B.3 gives a simplified visualization of applying a structural element. The image erosion idea is based on soil erosion, it erodes away the boundaries of the foreground object. Erosion removes white noise (i.e., signal power is independently distributed over time), but also shrinks our object, so we dilate it to regain the object area. In Chapter 3 of this thesis the morphological operations are used for joining disparate motion zones in the video sequence.

[1] `https://opencv.org/`

**Figure B.3** A structuring element is applied on the original image (left), the erosion operation (middle) makes the object becoming smaller and small objects disappear, the dilation operation (right) increases the object size and small objects are merged.


### B.1.5 Structural analysis and shape description

In the previous subsection, we discussed the use of a structuring element for noise removal. Similar filters can be used to extract useful information from an image, the features. Edge detection, the detection of sudden changes and discontinuities in the intensity function, for example is a simple technique to extract shape information from an image . Furthermore, the edges can be used to recognize the object or to identify the viewpoint and geometry of an image (see the visibility algorithm in Chapter 4). Subsequently, the features can also be used as input for the machine learning approach, which will be discussed in the following section.

The sudden change in discontinuity in the intensity profile can be visualized by the extrema of the derivative of that function. Furthermore the edge strength can be calculated by the gradient magnitude. It is important to remark that noise is highly affecting the filter response. Noise are pixels that have a very different intensity compared to their neighbors. The solution is to smooth the image before applying the convolution to find the derivative. Finally, this leads to different criteria to build an optimal edge detector .

- Good detection: minimize the false positives (noise classified as edge) and reduce the false negatives (missed real edges).
- Good localization: the detected edges should be very close to the real edges.
- Single response: only one detected point for each real edge point.

The method that best matches the previous criteria is the **canny edge detector [1]**. This frequently used edge detection algorithm consist of the following five steps:

- General image noise removal by applying a Gaussian filter,
- Gradient intensity retrieval: First, by applying a Sobel operator (e.g., a predefined 3x3 convolution mask) the 2D spatial gradient of the image is retrieved. Secondly, the direction and the magnitude of each edge is determined. Thirdly, the edge directions are reduced to four major directions (i.e., horizontal, vertical and the two diagonals).
- Non-maximum suppression to reduce the forged response to the edge detection and to merge all the weak detections that belong to the same edge.
- Content dependent double thresholding to remove false edge pixels due to noise and color variation.
- Edge tracking by hysteresis analysis (e.g., to connect short edge parts).

The edges are a basic feature descriptor that can be used as input in the machine learning approach (see Section B.2). It is important to remark that a descriptor should be: invariant with respect to the pose, scale, illumination of the object; highly distinctive in a large image dataset. In literature different methods are proposed (e.g., HOG: Histogram Of Gradients, SURF: Speeded Up Robust Features, Fast corner detector).

## B.2 Machine learning

The meaning of the term was already explained by Arthur Samuel in 1959 as: *Field of study that gives computers the ability to learn without being explicitly programmed.* The term resembles different techniques that identify patterns in observed data and build models that explain and predict new values. The different mechanisms are listed as follows:

- Association: a rule based method for discovering interesting relations (e.g., is there a link between people having a lower income and fire occurrence).

- Supervised learning: a methodology that infers a function or a model from labeled training data.

  - Classification; the output variable of the predictor is a category, a discrete value (e.g., fire and no-fire situation).
  - Regression: the output variable of the predictor is a real, continuous value (e.g., 80 percent visibility).

- Unsupervised learning: find insights in unlabeled information (e.g., cluster domestic house fires based on the average income)

- Reinforcement learning: solves the difficult problem of correlating immediate actions with the delayed returns they produce (e.g., learning an automatic fire extinguish robot by rewarding fire hits).

Besides the different models and mechanisms there is always a similar workflow that should be followed to ensure a stable system. Furthermore, the evaluation of a particular approach should be tackled decently. Within the following subsections we go more into detail in the workflow and the existing evaluation tools.

### B.2.1 Traditional approach

The traditional workflow consist of four major building blocks (i.e., firstly, the problem and data analysis, secondly the feature extraction or feature learning, thirdly the model selection and fourthly the model training and evaluation). It is important to remark the workflow should be repeated in a cycle manner to ensure a decent model that performs well during deployment.

- Problem and data analysis: this process involves a clear understanding of the data input and the expected output, the data resources. The first step is also to look into data cleaning (ensuring homogeneity) and to make the data manageable and accessible.

- Feature extraction: this can be done with 'hand crafted' features (e.g., shape representation, colors, edges) or with a pretrained network that looks for small representative patches.[2]
- Model selection and selecting the correct training approach. There are different models available, such as a Random Forest, a Support Vector Machine (SVM) or (Convolutional) Neural Network and each of them has its hyper-parameters that need to be optimized.
- Model training and evaluation: After training an initial model there is a continuous loop where we go back to the original data and problem. The performance of the model is evaluated via cross validation (e.g., a data splitting method to get a reliable estimate of the model performance by only looking to the training data) or via a separate test set. Depending on the task the evaluation and the selection of the best model can be done by looking to, for instance, the accuracy, recall, precision or the Mean Squared Error (MSE).

### B.2.2 Evaluation

Evaluation of a machine learning model is necessary to gain insights in the performance of the selected or trained model. Furthermore, the performance should be similar to the results that will be achieved during deployment. In the following sections we will shortly discuss the different types of evaluation metrics available in literature.

#### Bias versus variance

The bias error is the difference between the model predictions and the correct, real value. Whereas the variance is the error taken as the variability of the model predictions at a given point. There is a clear relationship between bias and the variance, if the bias is reduced the variance will increase whereas when the bias is increased the variance will decrease. This mechanism is in literature often called the Bias-Variance trade-off. Selecting a balance between the bias and variance error will lead to a system that minimizes overfitting and underfitting.

#### Accuracy, precision and recall

For the classification task (see Chapter 3 for a realistic example) there are different mechanisms available. The most frequently used evaluation metric is the accuracy. Still it is important to look into the cost of a misclassification for minor class samples (e.g., predicting no fire occurrence in case there is a real fire). Subsequently,

[2]Feature engineering refers to handcrafted features whereas feature learning refers to a system where it is automatically learns the representations directly from the data.

the combination of accuracy, precision, recall and F1-score will give more general insights in the misclassified samples.

The accuracy is the ratio between the correct predictions and the total number of predictions. The accuracy is a good indicator if the amount of samples in each class is equal (specific for classification)

$$Accuracy = \frac{TruePositive + FalseNegative}{TotalNumberOfSamples} \tag{B.4}$$

The precision is the ratio between the number of correct positive results and the total amount of positive results predicted by the model.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{B.5}$$

The recall is the ratio between the correct positive results and all the samples that should be classified as positive.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{B.6}$$

F1-score gives the balance between the precision and the recall and is a measure to know how precise and robust the classifier is.

$$F1 = 2 * \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{B.7}$$

### Cross-validation

Splitting the data into a training, validation and test set is necessary to optimize and evaluate the model properly. The training set is a sample of the data used to fit a particular model, the evaluation set provides an unbiased evaluation of the model fit while tuning the model hyperparameters and the test set allows to give an unbiased evaluation of the final model. Finally, it is important to have a balanced set of samples in each class. To create particular balanced sets you can use the cross-validation technique.

Cross-validation is a data splitting and model evaluation technique where you train the model on a subset of the input data and evaluate it on the complementary subset of the data. Subsequently, this mechanism gives insights in the generalization of the model and the overfitting behavior. K-fold cross-validation for instance is a frequently used technique where you split the data into k subsets, use k-1 subsets for training and based on the fold that was not used for training you perform the evaluation. Subsequently, you repeat this process k times and to measure the overall performance of the model you take the average of all the fold metrics.

**Overfitting**

Overfitting occurs when a model performs quite well on the training and evaluation datasources, but has failed to generalize on the test data. In Section B.3.4 we will go deeper into the overfitting handling for Convolutional neural networks.

## B.3 Convolutional neural networks

There are different machine learning models available, but the focus of this thesis lays on the visual and thermal image recognition. Based on previous studies CNN architectures are surpassing, SVM approaches for the recognition task. Therefore we only explain the basics of the CNN architecture. The workflow is similar to a standard neural network, which were inspired by the biological neural systems, were you learn the weights and the biases. Distinctive, the input for the CNN architecture is a 3D volume (width, height, depth (i.e., the color channels)). For more technical and mathematical details of the neural network architecture we refer to the following books [2, 3].

### B.3.1 Definitions

Before going into detail, different terms that are frequently used within a CNN model are explained. **Stride** is the number of pixels the filter kernel shifts over the input matrix on each learning step. **Padding** is the method of adding zeros (zero-padding) to fit the kernel perfectly with the input image. Another padding method is the removal of the borders where the kernel does not fit (valid-padding). **Rectified linear unit or ReLU** ensures that the output is positive also if the input is smaller or equal to zero. This results in faster training results for large networks. **Softmax** function results in a categorical probability distribution where the output is between 0 and 1, and the total sum of outputs is equal to 1.

### B.3.2 CNN architecture

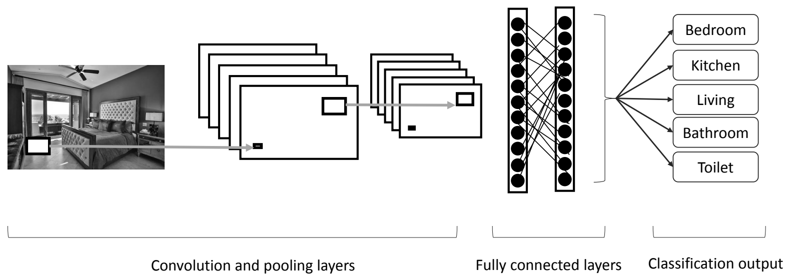
A simple CNN architecture consists of a minimal sequence of specific layers (convolutional, pooling and fully connected layers). Every layer transforms the volume of one filter response (activation) to another through a differentiable function. Figure B.4 gives a schematic overview of the combination of firstly the input image, secondly the convolutional and pooling layers, thirdly two fully connected layers and finally the classification output (the class label and the probability of the model).

**A convolutional layer** is the major building block of the CNN architecture. The filter is slided over the image spatially and computes the dot product of the filter and the original pixel or response (mathematical convolution).

**A pooling layer** is a layer that reduces the spatial size of the features. Furthermore, the layer reduces the amount of parameters and operates on each feature map. The most common form is a max pooling layer with a filter size of 2x2 and a stride of 2.

**A fully connected layer** is similar to a regular neural network where the neurons are connected to the entire input volume.

**Figure B.4** Basic CNN architecture for a classification task.




### B.3.3 Training procedure

The learning procedure of a convolutional neural network can be summarized as follows:

1. The weights of the filters (i.e., the layers) are initialized, this can be done randomly, via a specific distribution or according to the weights of a pre-trained network.

2. During the **forward pass** you input a training image and pass it though the whole network.

3. The output (the classification of the forwarded image) is compared to the expected value (the label) and **the loss** (i.e., the error between the expected and the real value) is calculated.

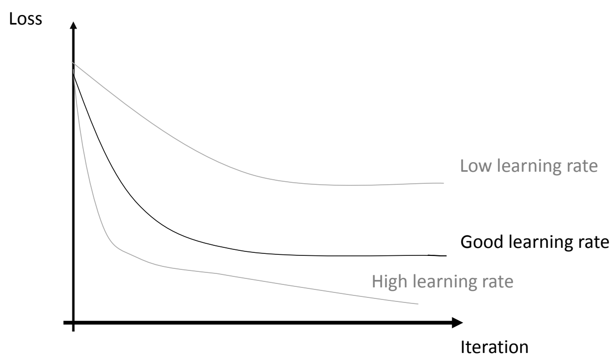
4. In the **backward pass** the error is propagated backwards through the CNN architecture and the impact of each weight on the loss is calculated (the derivation).

5. The **weights are updated** to reduce the loss.
6. The forward pass, the loss calculation, the backward pass and the weight updating are repeated for a fixed amount of iterations.

**Learning rate and the effect on the loss function**

The learning rate is one of the most important hyperparameters of the CNN model and based on the selected rate the loss function will be different. The rate controls how much we adjust the weighs of the filter with respect to the loss gradient. Figure B.5 visualizes the effect of different learning rates. If the learning rate is too high, the loss function will grow exponentially and divergence can be met. If the learning rate is too low the loss function will decrease very slow and possibly no convergence will be reached. Finally if the learning rate is appropriate the loss will decrease properly.

**Figure B.5** The impact of the learning rate on the loss function




**Weights and features**

Besides the standard evaluation metrics as proposed in SectionB.2.2 it is interesting to look in the visualization of the filter outputs (the weights). The low-level features (color, blobs, edges) are represented in the first filter activations whereas the high level features (object parts) are visualized in the higher activation maps.

### B.3.4 Overfitting and underfitting

The model is underfitting if the accuracy or the precision is higher on the validation set compared to the training set. Furthermore, the model will perform bad for the expected task.

To handle overfitting in CNN networks different mechanisms are proposed in literature. Most of the mechanisms will have a decent impact on the evaluation metrics and hyperparameter tuning will be necessary. A non limiting list of mechanisms is given below:

- Extend the dataset with more samples, this is in most cases not possible.
- Data augmentation or data creation from the original training dataset (e.g., flip, rotate, scale, zoom the images) to enlarge the original trainingset.
- Add regularization parameters:
  - Dropout: randomly discard a random set of activation functions during the training phase.
  - L1/ L2 regularization: changing the objective function that is minimized by adding a penalty for large weights.
- Reduce the architecture complexity: according to Ocam's Razor principle: select the solution or methodology with the fewest assumptions, the simplest method.

### B.3.5 Transfer learning

The major downside of CNN's is that they require a sufficient dataset size. Transfer learning is therefore an ideal technique where you 'transfer' the knowledge of the solution from a related task that has already been proven to work. Frequently, people pretrain the convolutional network on a large existing dataset (e.g., MNIST, CIFAR, Imagenet or COCO) and then use the network as a fixed feature extractor or use it for initialization.

The decision parameter to select a particular method is based on the size of the new dataset and on the similarity to the images of the pretrained network [4].

- **Convnet as a fixed feature extractor** From a pretrained convolutional network remove the last fully connected layer and use the remaining network as a fixed feature extractor. Subsequently, a linear classifier (i.e., a linear SVM or a small neural network) is trained on the extracted feature to perform the final classification. This method is affordable for small datasets with similar images to the original set where the model was trained on.

- **Finetuning a convnet with a small learning rate** Replace the last layer and fine-tune the weights of the filters with a small learning rate to avoid fundamental changes in the original trained model. This method is favorable in case the new dataset is large and similar to the original dataset. Subsequently, it is also possible to replace several layers and to train a linear classifier on activations (filter responses) earlier in the network (more generic features). This is useful if the new dataset is rather small, but different from the original dataset.

## B.4 Ensemble

In the previous sections we focused mainly on single models, however, combing different models (creating ensembles) can give improved results. Ensemble techniques [5] combine different models to make one final prediction, which is hopefully more reliant than any individual prediction. Ensemble techniques can be compared to the real life situation in which you ask various expert opinions (e.g., combine the output of object and scene detection as proposed in Chapter 3) to base your ultimate decision on. In literature different methods are proposed to combine the output of a particular model (i.e., stacking, bagging, boosting, plurality voting).

### B.4.1 Stacking

In this approach, the newly created model sees the individual predictions as features, and makes a final prediction. To avoid stimulating overfitting, stacking models have to be trained on individual predictions which were not included during the training of the individual models. Therefore, you can use the predictions on the 80-90 % and use the 90-100 % data interval to train/evaluate the stacking models. The main advantage of the stacking method is in the computational efficiency, and flexibility. By simply combining the individual predictions in one new feature matrix, only the ensemble need to be trained to make an ultimate prediction. Moreover, the use of output predictions allows to add new predictions at later stages and does not require the long computational times of consecutively training models.

### B.4.2 Plurality rule voting system

Plurality voting is a transparent method to combine different models to do a final prediction: simply select the class label that received the most votes. Often called majority voting, which is a special case of plurality voting for two class labels. Traditional plurality voting is appropriate as long as all classifiers are performing equally well. However in case the performance of the models is quite different it seems more appropriate to use the weighted variant of plurality voting: select the class label that has the largest number of weighted votes.

### B.4.3 Bagging

Another technique that can be used for ensemble creation is the Bootstrap aggregating Bagging technique in combination with Decision tree classifiers. In this case the model-trees are trained in parallel on bootstrap samples from the initial training set. The main goal of this approach is to decrease the variance and not the bias.

### B.4.4 Adaptive boosting

This sequential ensemble technique tries to add new models that perform well where the previous models lack. The main goal is to decrease the bias, still the model is sensitive to noisy data and outliers.

## References

[1] John Canny. *A computational approach to edge detection*. In Readings in Computer Vision, pages 184–203. Elsevier, 1987.

[2] Nasser M Nasrabadi. *Pattern recognition and machine learning*. Journal of electronic imaging, 16(4):049901, 2007.

[3] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[4] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. *How transferable are features in deep neural networks?* CoRR, abs/1411.1792, 2014.

[5] Thomas G Dietterich. *Ensemble learning*. The handbook of brain theory and neural networks, 2:110–125, 2002.